

CORRECTED  
VERSION\*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification 6 :</b> <u>A61K 38/39, 38/17, C12P 21/00</u>	<b>A1</b>	<b>(11) International Publication Number:</b> WO 97/38710 <b>(43) International Publication Date:</b> 23 October 1997 (23.10.97)
<b>(21) International Application Number:</b> PCT/US97/07300 <b>(22) International Filing Date:</b> 11 April 1997 (11.04.97) <b>(30) Priority Data:</b> 08/631,336 12 April 1996 (12.04.96) US <b>(60) Parent Application or Grant</b> <b>(63) Related by Continuation</b> US 08/631,336 (CIP) Filed on 12 April 1996 (12.04.96) <b>(71) Applicants (for all designated States except US):</b> FIBROGEN, INC. (US/US); 260 Littlefield Avenue, San Francisco, CA 94080 (US). ACADEMY OF FINLAND [FI/FI]; Hameenti 68B, FIN-00550 Helsinki (FI). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> KIVIRIKKI, Kari, I. [FI/FI]; Parkkisenranta 5, FIN-90650 Oulu (FI). PIHLA-JANIEMI, Taina [FI/FI]; Nauriskuja Runkotie 1B3, FIN-90460 Oulunsalo (FI). <b>(74) Agent:</b> HALLUIN, Albert, P.; Pennie & Edmonds L.L.P., 1155 Avenue of the Americas, New York, NY 10036 (US).	<b>(81) Designated States:</b> AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GE, GH, HU, IL, IS, JP, KG, KP, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TR, TT, UA, US, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
<b>(54) Title:</b> SYNTHESIS OF HUMAN PROCOLLAGENS AND COLLAGENS IN RECOMBINANT DNA SYSTEMS		
<b>(57) Abstract</b>  Methods of making collagen with hosts, and vectors that express collagen, and collagen post-translation enzymes are disclosed. Collagen post-translation enzymes include prolyl-4-hydroxylase, lysyl hydroxylase, lysyl oxidase, C-proteinase, and N-proteinase, and these enzymes increase the yield of properly folded, recombinant collagen in non-mammalian hosts. The collagens produced by these methods, hosts, and vectors include both homotrimer and heterotrimer collagen made from single or multiple collagen genes, respectively.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## SYNTHESIS OF HUMAN PROCOLLAGENS AND COLLAGENS IN RECOMBINANT DNA SYSTEMS

5

### 10 **RELATED APPLICATIONS**

This application is a continuation-in-part of United States Applications, Serial No. 08/631,336, filed April 12, 1996 ("the '336 Application"), which is a continuation-in-part of United States Applications, Serial No. 08/211,820, filed 15 August 11, 1994 ("the '820 Application"). The '820 Application is a U.S. National Application, pursuant to 35 U.S.C. § 371, of PCT Application Serial No. PCT/US92/09061, filed October 22, 1992, which is a continuation-in-part of U.S. Application No. 07/780,899, filed October 23, 1991, now abandoned. The '860 20 Application is a continuation-in-part of the '820 Application and United States Application Serial No. 08/210,063, filed March 16, 1994, which is a U.S. National Application, pursuant to 35 U.S.C. § 371, of PCT Application Serial No. PCT/US92/22333, filed June 10, 1992, which is a continuation of US Application Serial No. 07/713,945, filed June 12, 1991, now abandoned. Each of these 25 applications is incorporated herein by reference. Portions of the invention described herein were made in the course of research supported in part by NIH grants AR38188 and AR39740. The Government may have certain rights in this invention.

### 30 **I. FIELD OF THE INVENTION**

The present invention is directed to the recombinant production of procollagen, collagen and fragments thereof.

### 35 **II. BACKGROUND OF THE INVENTION**

*The ExtraCellular Matrix.* The most abundant component of the extracellular matrix is collagen. Collagen molecules are generally the result of the trimeric

assembly of three polypeptide chains containing, in their primary sequence, (-Gly-X-Y-)n repeats which allow for the formation of triple helical domains (van der Rest *et al.* FASEB J. 5:2814-2823 (1991)). During their biosynthesis, the three polypeptide chains comprising collagen undergo various post-translational modifications which permit the formation of these triple helical domains (Van der Rest *et al.*, Adv. Mol. Cell Biol. 6:1-67 (1993)). For example, the proline residues of collagen are hydroxylated into 4- hydroxyproline, thereby allowing for the formation of interchain hydrogen bonds by the enzyme prolyl 4-hydroxylase (Kivirikko *et al.*, Post-translational modifications of proteins (Harding, J. J., Crabbe, M. J. C., eds) pp. 1-51, CRC Press, Boca Raton, FL (1992)). The triple-helical molecule is then further processed to render collagens. For example, the N-propeptide and C-propeptide comprising the collagen precursor molecule, "procollagen," are cleaved during post-translational events by the enzymes N-proteinase and C-proteinase, respectively.

As a consequence of the diverse structural and functional properties of collagen in its various forms or "types," collagen can contribute significantly to the high diversity of the extracellular matrix.

*Collagen Types.* Nineteen distinct collagen types have been identified in vertebrates, including bovine, ovine, porcine, chicken and human collagens. These collagen types are numbered by Roman numerals and the chains found in each collagen type are identified with Arabic numerals. A detailed description of structure and biological functions of the various different types of naturally occurring collagens can be found, among other places, in Ayad *et al.*, The Extracellular Matrix Facts Book, Academic Press, San Diego, CA; Burgeson, R. E., and Nimmi, "Collagen types: Molecular Structure and Tissue Distribution," Clin. Orthop. 282:250-272 (1992); Kielty, C. M. *et al.*, "The Collagen Family: Structure, Assembly And Organization In The Extracellular Matrix," in Connective Tissue And Its Heritable Disorders, Molecular Genetics, And Medical Aspects, Royce, P. M. and Steinmann, B., Eds., Wiley-Liss, NY, pp. 103-147 (1993).

Type I collagen is the major fibrillar collagen of bone and skin. Type I collagen is a heterotrimeric molecule comprising two  $\alpha 1(I)$  chains and one  $\alpha 2(I)$

chain. Details on preparing purified type I collagen can be found, among other places, in Miller *et al.*, Methods In Enzymology 82:33-64 (1982), Academic Press.

Type II collagen is a homotrimeric collagen comprising three identical  $\alpha$ (II) chains. Purified Type II collagen may be prepared from tissues by, among other methods, the procedure described in Miller *et al.*, Methods In Enzymology, 82:33-64 (1982), Academic Press.

Type III collagen is a major fibrillar collagen found in skin and vascular tissues. Type III collagen is a homotrimeric collagen comprising three identical  $\alpha$ (III) chains. Methods for purifying type III collagen from tissues can be found in, among other places, Byers *et al.*, Biochemistry 13:5243-5248 (1974) and Miller *et al.*, Methods in Enzymology 82:33-64 (1982), Academic Press.

Type IV collagen is found in basement membranes in the form of a sheet rather than fibrils. The most common form of type IV collagen contains two  $\alpha$ 1(IV) chains and one  $\alpha$ 2(IV) chain. The particular chains comprising type IV collagen are tissue-specific. Type IV collagen may be purified by, among other methods, the procedures described in Furuto *et al.*, Methods in Enzymology 144:41-61 (1987), Academic Press.

Type V collagen is a fibrillar collagen found in, primarily, bones, tendon, cornea, skin, and blood vessels. Type V collagen exists in both homotrimeric and heterotrimeric forms. One type of type V collagen is a heterotrimer of two  $\alpha$ 1(V) chains and  $\alpha$ 2(V). Another type of type V collagen is a heterotrimer of  $\alpha$ 1(V),  $\alpha$ 2(V), and  $\alpha$ 3(V). Yet another type of type V collagen is a homotrimer of  $\alpha$ 1(V). Methods for isolating type V collagen from natural sources can be found, among other places, in Elstrow *et al.*, Collagen Rel. Res. 3:181-193 (1983) and Abedin *et al.*, Biosci. Rep. 2:493-502 (1982).

Type VI collagen has a small triple helical region and two large non-collagenous remainder portions. Type VI collagen is a heterotrimer comprising  $\alpha$ 1(VI),  $\alpha$ 2(VI), and  $\alpha$ 3(VI) chains. Type VI collagen is found in many connective tissues. Descriptions of how to purify type VI collagen from natural sources can be found, among other places, in Wu *et al.*, Biochem. J. 248:373-381 (1987), and Kielty *et al.*, J. Cell Sci. 99:797-807.

Type VII collagen is a fibrillar collagen found in particular epithelial tissues. Type VII is a homotrimeric molecule of three  $\alpha 1(\text{VII})$  chains. Descriptions of how to purify type VII collagen from tissue can be found in, among other places, 5 Lundstrom *et al.*, J. Biol. Chem. **261**:9042-9048 (1986), and Bentz *et al.*, Proc. Natl. Acad. Sci. USA **80**:3168-3172 (1983).

Type VIII collagen can be found in Descemet's membrane in the cornea. Type VIII collagen is a heterotrimer comprising two  $\alpha 1(\text{VIII})$  chains and one  $\alpha 2(\text{VIII})$  chain, although other chain compositions have been reported. Methods for 10 the purification of type VIII collagen from nature can be found, among other places, in Benya *et al.*, J. Biol. Chem. **261**:4160-4169 (1986), and Kapoor *et al.*, Biochemistry **25**:3930-3937 (1986).

Type IX collagen is a fibril associated collagen which can be found in 15 cartilage and vitreous humor. Type IX collagen is a heterotrimeric molecule comprising  $\alpha 1(\text{IX})$ ,  $\alpha 2(\text{IX})$ , and  $\alpha 3(\text{IX})$  chains. Procedures for purifying type IX collagen can be found, among other places, in Duance *et al.*, Biochem. J. **221**:885-889 (1984), Ayad *et al.*, Biochem. J. **262**:753-761 (1989), Grant *et al.*, 20 The Control of Tissue Damage, Glauert, A. M., Ed., El Sevier, Amsterdam, pp. 3-28 (1988).

Type X collagen is a homotrimeric compound of  $\alpha 1(\text{X})$  chains. Type X collagen has been isolated from, among other tissues, hypertrophic cartilage found in growth plates. 25

Type XI collagen can be found in cartilaginous tissues associated with type II and type IX collagens, as well as other locations in the body. Type XI collagen is a heterotrimeric molecule comprising  $\alpha 1(\text{XI})$ ,  $\alpha 2(\text{XI})$ , and  $\alpha 3(\text{XI})$  chains. Methods for purifying type XI collagen can be found, among other places, in Grant *et al.*, In The 30 Control of Tissue Damage, Glauert, A. M., ed., El Savier, Amsterdam, pp.3-28 (1988).

Type XII collagen is a fibril associated collagen found primarily associated with type I collagen. Type XII collagen is a homotrimeric molecule comprising 35 three  $\alpha 1(\text{XII})$  chains. Methods for purifying type XII collagen and variants thereof can be found, among other places, in Dublet *et al.*, J. Biol. Chem. **264**:13150-13156

(1989), Lundstrum *et al.*, J. Biol. Chem. 267:20087-20092 (1992). Watt *et al.*, J. Biol. Chem. 267:20093-20099 (1992).

5 Type XIII is a non-fibrillar collagen found, among other places, in skin, intestine, bone, cartilage, and striated muscle. A detailed description of the type XIII collagen may be found, among other places, in Juvonen *et al.* J. Biol. Chem. 267:24700-24707 (1992).

10 Type XIV is a fibril associated collagen. Type XIV collagen is a homotrimeric molecule comprising three  $\alpha 1$ (XIV) chains. Methods for isolating type XIV collagen can be found, among other places, in Aubert-Foucher *et al.*, J. Biol. Chem. 266:19759-19764 (1992) and Watt *et al.*, J. Biol. Chem. 267:20093-20099 (1992).

15 Type XV collagen is homologous in structure to type XVIII collagen. Information about the structure and isolation of natural type XV collagen can be found, among other places, in Myers *et al.*, Proc. Natl. Acad. Sci. USA 89:10144-10148 (1992), Huebner *et al.*, Genomics 14:220-224 (1992), Kivirikko *et al.*, J. Biol. Chem. 269:4773-4779 (1994), and Muragaki, J. Biol. Chem. 264:4042-4046 (1994).

20 Type XVI collagen is a fibril associated collagen, found in skin, lung fibroblast, keratinocytes, and elsewhere. Information on the structure of type XVI collagen and the gene encoding type XVI can be found, among elsewhere, in Pan *et al.*, Proc. Natl. Acad. Sci. USA 1989:6565-6569 (1992), and Yamaguchi *et al.*, J. Biochem. 112:856-863 (1992).

25 Type XVII collagen is a hemidesmosomal transmembrane collagen. Information on the structure of type XVII collagen and the gene encoding type XVII collagen can be found, among elsewhere, in Li *et al.*, J. Biol. Chem. 268(12):8825-8834 (1993), and McGrath *et al.*, Nat. Genet. 11(1):83-86 (1995).

30 Type XVIII collagen is similar in structure to type XV collagen and can be isolated from the liver. Descriptions of the structures and isolation of type XVIII collagen from natural sources can be found, among other places, in Rehn *et al.*, Proc. Natl. Acad. Sci. USA 91:4234-4238 (1994), Oh *et al.*, Proc. Natl. Acad. Sci.

USA 91:4229- 4233 (1994), Rehn *et al.*, J. Biol. Chem. 269:13924-13935 (1994), and Oh *et al.*, Genomics 19:994-999 (1994).

Type XIX collagen's gene structure classify it as another member of the FACIT collagenous family. Type XIX mRNA was recently isolated from  
5 rhabdomyosarcoma cell. Descriptions of the structures and isolation of type XIX collagen can be found, among other places, in Inoguchi *et al.*, J. Biochem. 117:137-146 (1995), Yoshioka *et al.*, Genomics 13:884-886 (1992), Myers *et al.*, J. Biol. Chem. 269:18549-18557 (1994).

10 **Post-Translational Enzymes.** Prolyl 4-hydroxylase is an important post-translational enzyme necessary for the synthesis of procollagen or collagen by cells. The enzyme is required to hydroxylate prolyl residues in the Y-position of the repeating -Gly-X-Y- sequences to 4-hydroxyproline. Prockop *et al.*, N. Engl. J.  
15 Med. 311:376-386 (1984). Unless an appropriate number of Y-position prolyl residues are hydroxylated to 4-hydroxyproline by prolyl 4- hydroxylase, the newly synthesized chains cannot fold into a triple-helical conformation at 37°C. Moreover, if the hydroxylation does not occur, the polypeptides remain non-helical, are poorly  
20 secreted by cells, and cannot self- assemble into collagen fibrils.

Prolyl-4-hydroxylase from vertebrates is an  $\alpha_2\beta_2$  tetramer. Berg *et al.*, J. Biol. Chem. 248:1175-1192 (1973); Tuderman *et al.*, Eur. J. Biochem. 52:9-16 (1975). The  $\alpha$  subunits (~63 kDa) contain the catalytic sites involved in the hydroxylation of prolyl residues but are insoluble in the absence of  $\beta$  subunits. The  
25  $\beta$  subunits (~55 kDa) were found to be identical to the protein disulfide isomerase, which catalyzes thiol/disulfide interchange in a protein substrate, leading to the formation of the set of disulfide bonds which permit establishment of the most stable state of the protein. The  $\beta$  subunits retain 50% of protein disulfide isomerase  
30 activity when part of the prolyl-4-hydroxylase tetramer. Pihlajaniemi *et al.*, Embo J. 6:643-649 (1987); Parkkonen *et al.*, Biochem. J. 256:1005-1011 (1988); Koivu *et al.*, J. Biol. Chem. 262:6447-6449 (1987)). Recently, active recombinant human enzyme has been produced in insect cells by simultaneously expressing the  $\alpha$  and  $\beta$   
35 subunits in Sf9 cells. Vuori *et al.*, Proc. Natl. Acad. Sci. USA 89:7467-7470 (1992).



In addition to prolyl-4-hydroxylase, other collagen post-translational enzymes have been identified and reported in the literature, including C-proteinase, N-proteinase, lysyl oxidase, and lysyl hydroxylase.

5       ***Attempts to Express Collagen.*** Expression of many exogenous genes is readily obtained in a variety of recombinant host-vector systems. Expression, however, becomes difficult to obtain if the final formation of the protein requires extensive post-translational processing. This is the likely reason that, prior to the present invention, expression of properly formed collagen in a fully recombinant  
10 system has not been reported. See Prockop *et al.*, N. Engl. J. Med. 311:376-386 (1984).

Notably, rescue experiments in two different systems that synthesized only one of the two chains for type I procollagen have been reported. Specifically, it was  
15 found that a gene for the human fibrillar procollagen pro $\alpha$ 1(I) chain, the COL1A1 gene, can be expressed in mouse fibroblasts and the chains used to assemble molecules of type I procollagen, the precursor of type I collagen. However, the reports are limited to the pro $\alpha$ 2(I) chains of mouse origin. Hence, the type I  
20 procollagen synthesized is a hybrid molecule of human and mouse origin.

Similarly, expression of a rat exogenous pro $\alpha$ 2(I) gene to generate type I rat procollagen have been reported. Thus, synthesis of a recombinant procollagen molecule in which all three chains are derived from exogenous genes was not obtained in the art.

25       Failure to obtain expression of genes for human collagens has made it impossible to prepare human procollagens and collagens that have a number of therapeutic uses in man and that will not produce the undesirable immune responses that have been encountered with use of collagen from animal sources. Also, many  
30 types of collagen are only available in trace quantities in tissues and can only be obtained in significant quantities by recombinant production.

### III. SUMMARY OF THE INVENTION

**Methods.** The present invention comprises the expression of at least one nucleic acid sequence encoding a collagen chain, and at least one nucleic acid sequence encoding a collagen post-translational enzyme.

5 More specifically, the present invention provides for methods of expressing at least a single procollagen or collagen gene (or other nucleic acid molecule) or a number of different procollagen or collagen genes (or other nucleic acid molecule) within a cell. Further, it is contemplated that there can be one or more copies of a  
10 single procollagen or collagen gene (or other nucleic acid molecule) or of the number of different such genes introduced into cells (*i.e.*, transformation or transduction) and expressed. The present invention provides that these cells can be transformed or transfected with nucleic acids encoding collagen and enzymes that modify collagen so  
15 that they express at least one procollagen or collagen chain, preferably human, that will assemble into a homotrimer or heterotrimer procollagen or collagen.

In one embodiment of the present invention, the method utilizes a procollagen or collagen gene (or other nucleic acid molecule) transfected into and expressed  
20 within cells which are a mutant, variant, hybrid or recombinant gene (or other nucleic acid molecule). Such mutant, variant, hybrid or recombinant gene may include, for example, a mutation which provides unique restriction sites for cleavage of the hybrid gene.

In a further embodiment of the present invention, such mutations provide one  
25 or more unique restriction sites do not alter the amino acid sequence encoded by the nucleic acid molecule, but merely provide unique restriction sites useful for manipulation of the molecule. Thus, the modified molecule would be made up of a number of discrete regions, or D-regions, flanked by unique restriction sites. These  
30 discrete regions of the molecule are herein referred to as cassettes. For example, cassettes designated as D1 through D4.4 are shown in Figure 4. Molecules formed of multiple copies of a cassette are another variant of the present gene which is encompassed by the present invention. Recombinant or mutant nucleic acid  
35 molecules or cassettes which provide desired characteristics such as resistance to

endogenous enzymes such as collagenase are also encompassed by the present invention.

A novel feature of the methods of the invention is that relatively large amounts of a human procollagen or collagen can be synthesized in a recombinant cell culture system that does not make any other procollagen or collagen. Systems that make other procollagens or collagens are preferred because of the extreme difficulty of separating the product of the endogenous genes for procollagen or collagen from recombinant collagen products. Using methods of the present invention, purification of procollagen, including human, bovine, porcine, chicken and other mammalian collagens, is greatly facilitated. Moreover, it has been demonstrated that the amounts of protein synthesized by the methods of the present invention are high relative to other systems used in the art.

Other novel features of the methods of the present invention are that procollagens synthesized are correctly folded proteins so that they exhibit the normal triple-helical conformation characteristic of procollagens and collagens. Therefore, the procollagens can be used to generate stable collagen by cleavage of the procollagens with proteases.

The present invention provides methods for the production of procollagens or collagens derived solely from transformed or transfected procollagen and collagen genes, such methods are not limited, however, to the production of procollagen and collagen derived solely from transformed or transfected genes.

**Vectors.** The present invention is also directed to vectors and plasmids used in the methods of the invention. Such vectors and/or plasmids are comprised of the nucleic acid sequence encoding the desired procollagens and collagens and necessary promoters, and other sequences necessary for the proper expression of such procollagens and collagens. In a preferred embodiment, the vectors and plasmids of the present invention further include at least one sequence encoding one or more post-translational enzymes.

It is an object of the invention to construct expression vectors for various host cells that contain collagen genes from human and other sources, and to construct

expression vectors that contain various collagen post- translation modification enzymes.

**Cells.** The present invention further comprises cells in which a procollagen or collagen, either alone or in combination with one or more post translational  
5 enzymes, is expressed both as mRNA and as a protein. Preferably, the procollagen or collagen (types I-XIX), and/or the post-translational enzyme, is expressed in mammalian cells, insect cells, or yeast cells. Notwithstanding these preferred embodiments, other cells, including plant cells and algae, can be manufactured.

10 In preferred embodiments of the present invention, cells such as mammalian, insect and yeast cells, which may not naturally produce sufficient amounts of post-translational enzymes, are transformed with at least one set of genes coding for a post-translational enzyme, such as prolyl 4-hydroxylase, C-proteinase, N-proteinase,  
15 lysyl oxidase or lysyl hydroxylase.

**Polypeptides.** The invention comprises the recombinant polypeptides expressed according to the methods of the present invention, including fusion products produced from chimeric genes wherein, for example, relevant epitopes of  
20 collagen or procollagen can be manufactured for therapeutic and other uses. The polypeptides of the present invention further include deglycosolated, unglycosolated and partially glycosolated collagens and procollagens.

An advantage of recombinant collagens of the present invention is that these collagens will not produce allergic responses in the mammals to which they are  
25 administered provided that the recombinant collagen is manufactured utilizing the nucleic acid sequence encoding such mammal's native collagen . For example, it is expected that humans will be to tolerate the administration of human collagen, as compared to collagen derived from other mammals (*e.g.*, bovine derived collagen).  
30 Moreover, collagen of the present invention prepared from cultured cells should be of a higher quality than collagen obtained from animal sources, and should form larger and more tightly packed proteins.

35

#### IV. BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a photograph showing analysis by polyacrylamide gel electrophoresis in SDS of the proteins secreted into medium by HT-1080 cells that were transfected with a gene construct containing the promoter, first exon and most of the first intron of the human COL1A1 gene linked to 30 kb fragment containing all of COL2A1 except the first two exons.

Figure 2 is a photograph evidencing the secretion type II procollagen into the medium from cells described in Figure 1 was folded into a correct native conformation.

Figure 3 is a photograph showing analysis of medium of HT-1080 cells co-transfected with a gene for COL1A1 and a gene for COL1A2.

Figure 4 is a schematic representation of the cDNA for the pro $\alpha$ 1(I) chain of human type I procollagen that has been modified to contain artificial sites for cleavage by specific restriction endonucleases.

Figure 5 is a photograph showing analysis by nondenaturing 7.5% polyacrylamide gel electrophoresis (lanes 1-3) and 10% polyacrylamide gel electrophoresis in SDS (lanes 4-6) of purified chick prolyl 4-hydroxylase (lanes 1 and 4) and the proteins secreted into medium by Sf9 cells expressing the gene for the  $\alpha$ -subunit and the  $\beta$ -subunit of human prolyl 4-hydroxylase and infected with  $\alpha$ 58/B virus (lanes 2 and 5) or with  $\alpha$ 59/B virus (lanes 3 and 6).  $\alpha$ 58/B and  $\alpha$ 59/B differ by a stretch of 64 base pairs.

Figure 6 is a gel showing the expression of recombinant human type III procollagen in Sf9 and High Five cells.

Figure 7 is a gel showing the expression of recombinant human type I procollagen in insect cells, analyzed on a silver stained, 5% SDS-PAGE gel. Lane 1 is a pepsin digested sample from cells expressing only the pro  $\alpha$ 1 chain of type I procollagen. Lane 2 is a pepsin digested sample from cells coexpressing pro $\alpha$ 1 and pro $\alpha$ 2 chains of type I procollagen.

Figure 8 is a gel showing the expression of recombinant human type II procollagen in insect cells, analyzed on a coomassie stained 5% SDS-PAGE gel.

Figure 9 is an SDS-PAGE analysis under reducing and nonreducing conditions of purified type III collagen. The gel was stained with Coomassie Brilliant Blue. The reduced type III collagen sample is shown in lane 2 and the nonreduced sample in lane 3. Molecular weight markers were run in lane 1. The positions of the trimeric  $\alpha 1$  (III) chains and the monomeric  $\alpha 1$  (III) chains are shown by arrows.

Figure 10 is a non-reducing SDS-PAGE analysis of trimer formation of the pro $\alpha 1$  (III) chains expressed in High Five insect cells. The samples were electrophoresed on 5% SDS-PAGE under nonreducing conditions and analyzed by Coomassie staining. Lane 1, molecular weight markers; lane 2, cell extract; lane 3, cell extract digested with pepsin; lane 4, proteins soluble in 1% SDS. The positions of the trimeric pro $\alpha 1$  (III) and  $\alpha 1$  (III) chains are shown by arrows.

Figures 11A-11D is an analysis of the thermal stability of the recombinant human type III collagen produced in insect cells by a brief protease digestion.

## V. DETAILED DESCRIPTION OF THE INVENTION

### A. Definitions:

The term "collagen" refers to any one of the collagen types I-XIX, as well as any novel collagens produced according to the methods of this invention. The term also encompasses both procollagen and mature collagen assembled as hetero- and homo-trimers, and any single chain polypeptides of procollagen or collagen for any of the collagen types, and any heterotrimers of any combination of the collagen constructs of the invention. The term "collagen" is meant to encompass all of the foregoing, unless the context dictates otherwise.

The term "procollagen" refers to any one of the collagen types I-XIX, as well as any novel collagens produced by this invention, that possess additional C-terminal and/or N-terminal peptides that assist in trimer assembly, solubility, purification or other function, and then are subsequently cleaved by N-proteinase, C-proteinase or other proteins.

The term "collagen subunit" refers to the amino acid sequence of one polypeptide chain of a collagen protein encoded by a single gene, as well as derivatives, including deletion derivatives, conservative substitutions, etc.

5 A "fusion protein" is a protein in which peptide sequences from different proteins are covalently linked together.

The term "collagen post-translational enzyme" refers to any enzyme that modifies a procollagen, collagen, or components comprising a collagen molecule, and encompasses, but is not limited to, prolyl-4-hydroxylase, C-proteinase,  
10 N-proteinase, lysyl hydroxylase, and lysyl oxidase. The term "collagen post-translational enzyme" is meant to encompass all of the foregoing, unless the context dictates otherwise.

The term "infection" refers to the introduction of nucleic acids into an  
15 organism by use of a virus or viral vector, and preferably, baculovirus or Semliki Forest virus.

The term "transformation" means introducing DNA into an organism so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal  
20 integration.

The term "transfection" refers to the taking up of an expression vector by a host cell, whether or not any coding sequences are in fact expressed.

The phrase "stringent conditions" as used herein refers to those hybridizing conditions that (1) employ low ionic strength and high temperature for washing, for  
25 example, 0.015 M NaCl/0.0015 M sodium citrate/0.1% SDS at 50°C.; or (2) employ during hybridization a denaturing agent such as formamide, for example, 50% (vol/vol) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl,  
30 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M Sodium citrate), 5 x Denhardt's solution, sonicated salmon sperm DNA (50 g/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC and 0.1% SDS.

35 The term "purified" as used herein denotes that the indicated collagen or procollagen is present in the substantial absence of other biological macromolecules,

e.g., polynucleotides, proteins, and the like. The term "purified" as used herein preferably means at least 95% by weight, more preferably at least 99.8% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a protein molecule separated not only from other proteins that are present in the natural source of the protein, but also from other proteins, and preferably refers to a protein found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass proteins present in their natural source.

#### 15        B.        Nucleic Acids Related To The Present Invention

In accordance with the invention, polynucleotide sequences which encode any collagen subunit, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that subunit of collagen, or a functional equivalent thereof, in appropriate host cells. Preferred 20 embodiments of the invention relate to polynucleotide sequences encoding human collagens or functional equivalents thereof. Preferred embodiments of the invention also include the polynucleotide sequences of collagen subunits of type I - type IV, type XIII, type XV, and type XVIII, or functional equivalents thereof.

25        The nucleic acid sequences encoding the known collagen types have been generally described in the art. See, e.g., Fukai *et al.*, Methods of Enzymology 245:3-28 (1994) and references cited therein. New collagens/procollagens or known collagens/procollagens from which nucleic acid sequence is not available may be 30 obtained from cDNA libraries prepared from tissues believed to possess a "novel" type of collagen and to express the novel collagen at a detectable level. For example, a cDNA library could be constructed by obtaining polyadenylated mRNA from a cell line known to express the novel collagen, or a cDNA library previously 35 made to the tissue/cell type could be used. The cDNA library is screened with appropriate nucleic acid probes, and/or the library is screened with suitable



polyclonal or monoclonal antibodies that specifically recognize other collagens.

Appropriate nucleic acid probes include oligonucleotide probes that encode known portions of the novel collagen from the same or different species. Other suitable probes include, without limitation, oligonucleotides, cDNAs, or fragments thereof that encode the same or similar gene, and/or homologous genomic DNAs or fragments thereof. Screening the cDNA or genomic library with the selected probe may be accomplished using standard procedures known to those in the art, such as those described in Chapters 10-12 of Sambrook *et al.*, Molecular Cloning: A Laboratory Manual. New York, Cold Spring Harbor Laboratory Press, 1989. Other means for identifying novel collagens involve known techniques of recombinant DNA technology, such as by direct expression cloning or using the polymerase chain reaction (PCR) as described in U.S. Patent No. 4,683,195, issued 28 July 1987, or in section 14 of Sambrook *et al.*, Molecular Cloning: A Laboratory Manual. Second Edition, Cold Spring Harbor Laboratory Press, New York, 1989, or in Chapter 15 of Current Protocols in Molecular Biology, Ausubel *et al.* eds., Greene Publishing Associates and Wiley-Interscience 1991.

Altered DNA sequences which may be used in accordance with the invention include deletions, additions or substitutions of different nucleotide residues resulting in a sequence that encodes the same or a functionally equivalent gene product. The gene product itself may contain deletions, additions or substitutions of amino acid residues within a collagen sequence, which result in a functionally equivalent collagen.

The nucleic acid sequences of the invention may be engineered in order to alter the collagen coding sequence for a variety of ends including, but not limited to, alterations which modify processing and expression of the gene product. For example, alternative secretory signals may be substituted for the native human secretory signal and/or mutations may be introduced using techniques which are well known in the art, *e.g.*, site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, phosphorylation, etc. Additionally, when expressing in non-human cells, the polynucleotides encoding the collagens of the invention may be

modified in the silent position of any triplet amino acid codon so as to better conform to the codon preference of the particular host organism.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described collagens and fragments. These amino acid sequence variants of native collagens and collagen fragments may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant collagen encoding polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. The amino acid sequence variants of collagen are preferably constructed by mutating the polynucleotide to give an amino acid sequence that does not occur in nature. These amino acid alterations can be made at sites that differ in collagens from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site.

Amino acids are divided into groups based on the properties of their side chains (polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipatic nature): (1) hydrophobic (leu, met, ala, ile), (2) neutral hydrophobic (cys, ser, thr), (3) acidic (asp, glu), (4) weakly basic (asn, gln, his), (5) strongly basic (lys, arg), (6) residues that influence chain orientation (gly, pro), and (7) aromatic (trp, tyr, phe). Conservative changes encompass variants of an amino acid position that are within the same group as the "native" amino acid. Moderately conservative changes encompass variants of an amino acid position that are in a group that is closely related to the "native" amino acid (*e.g.*, neutral hydrophobic to weakly basic). Non-conservative changes encompass variants of an amino acid position that are in a group that is distantly related to the "native" amino acid (*e.g.*, hydrophobic to strongly basic or acidic).

Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid

insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal  
5 insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells.

In a preferred method, polynucleotides encoding a collagen are changed via site-directed mutagenesis. This method uses oligonucleotide sequences that encode  
10 the polynucleotide sequence of the desired amino acid variant, as well as a sufficient adjacent nucleotide on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is  
15 exemplified by publications such as, Edelman *et al.*, DNA 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, Nucleic Acids Res. 10:6487-6500 (1982).

PCR may also be used to create amino acid sequence variants of a collagen.  
20 When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the collagen at the position specified by the primer. The product DNA fragments  
25 replace the corresponding region in the plasmid and this gives the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells *et al.*, Gene 34:315 (1985); and other  
30 mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook *et al.*, supra, and Current Protocols in Molecular Biology, Ausubel *et al.*, supra.

In another embodiment of the invention, a collagen sequence may be ligated  
35 to a heterologous sequence to encode a fusion protein. For example, a fusion protein may be engineered to contain a cleavage site located between an (3IX) collagen

sequence and the heterologous protein sequence, so that the (3IX) collagen may be cleaved away from the heterologous moiety.

Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these collagen proteins. Such DNA sequences include those which are capable of hybridizing to the appropriate human collagen sequence under stringent conditions.

10 C. Collagen Modifying Polypeptides And Corresponding Nucleic Acid Sequences

As naturally produced, collagens are structural proteins comprised of one or more collagen subunits which together form at least one triple-helical domain. A variety of enzymes are utilized in order to transform the collagen subunits into procollagen or other precursor molecules and then mature collagen. Such enzymes include prolyl-4-hydroxylase, C-proteinase, N-proteinase, lysyl oxidase and lysyl hydroxylase.

Prolyl 4-hydroxylase plays a central role in the biosynthesis of all collagens, as the 4-hydroxyproline residues stabilize the folding of the newly synthesized polypeptide chains, into triple-helical molecules. Prockop *et al.*, Annu. Rev. Biochem. 64:403-434 (1995); Kivirikko *et al.*, "Post-Translational Modifications of Proteins," pp. 1-51 (1992); Kivirikko *et al.*, FASEB J. 3:1609-1617 (1989). For example, when the pro $\alpha$ 1 chains of type III procollagen were expressed in insect cells, without recombinant prolyl 4-hydroxylase, considerable amounts of procollagen were made in the cells, and the pro $\alpha$ 1 chains formed triple-helical molecules as indicated by the resistance of the collagenous domains of the collagen to protease degradation at 22°C. However, the T<sub>m</sub> of the triple helices of such molecules was about 6°C lower than procollagen produced in the presence of the recombinant prolyl 4-hydroxylase. Also, the level of expression of type III collagen was lower in the absence of recombinant prolyl 4-hydroxylase than in its presence.

Lysyl hydroxylase, an  $\alpha$ 2 homodimer, catalyzes the post-translation modification of collagen to form hydroxylysine in collagens. See generally, Kivirikko *et al.*, Post-Translational Modifications of Proteins, Harding, J.J., and

Crabbe, M.J.C., eds., CRC Press, Boca Raton, FL (1992); Kivirikko, Principles of Medical Biology, Vol. 3 Cellular Organelles and the Extracellular Matrix, Bittar, E.E., and Bittar, N., eds., JAI Press, Greenwich, Great Britain (1995).

5 C-proteinase processes the assembled procollagen by cleaving off the C-terminal ends of the procollagens that assist in assembly of, but are not part of, the triple helix of the collagen molecule. See generally, Kadler *et al.*, J. Biol. Chem. 262:15969-15701 (1987), Kadler *et al.*, Ann. NY Acad. Sci. 580:214-224 (1990).

10 N-proteinase processes the assembled procollagen by cleaving off the N-terminal ends of the procollagens that assist in the assembly of, but are not part of, the collagen triple helix. See generally, Hojima *et al.*, J. Biol. Chem. 269:11381-11390 (1994).

Lysyl oxidase is an extracellular copper enzyme that catalyzes the oxidative  
15 deamination of the  $\epsilon$ -amino group in certain lysine and hydroxy lysine residues to form a reactive aldehyde. These aldehydes then undergo an aldol condensation to form aldols, which cross links collagen fibrils. Information on the DNA and protein sequence of lysyl oxidase can found, among elsewhere, in Kivirikko, Principles of  
20 Medical Biology, Vol. 3 Cellular Organelles and the Extracellular Matrix, Bittar, E.E., and Bittar, N., eds., JAI Press, Greenwich, Great Britain (1995), Kagan, Path. Res. Pract. 190: 910-919 (1994), Kenyon *et al.*, J. Biol. Chem. 268(25):18435-18437 (1993), Wu *et al.*, J. Biol. Chem. 267(34):24199-24206 (1992), Mariani *et al.*, Matrix 12(3):242-248 (1992), and Hamalainen *et al.*,  
25 Genomics 11(3):508-516 (1991).

The nucleic acid sequences encoding a number of these post-translational enzymes have been reported. See e.g. Vuori *et al.*, Proc. Natl. Acad. Sci. USA 89:7467-7470 (1992); Kessler *et al.*, Science 271:360-362 (1996). The nucleic acid  
30 sequences encoding the various post-translational enzymes may also be determined according to the methods generally described above and include use of appropriate probes and nucleic acid libraries.

**D. Host-Vector Systems for Expressing Recombinant Collagen**

In order to express the collagens and related collagen post-translational enzymes of the invention, the nucleotide sequence encoding the collagen, or a functional equivalent, is inserted into an appropriate expression vector, *i.e.*, a vector  
5 which contains the necessary elements for the transcription and translation of the inserted coding sequence, or in the case of an RNA viral vector, the necessary elements for replication and translation.

Methods which are well known to those skilled in the art can be used to  
10 construct expression vectors containing a collagen coding sequence for the collagens of the invention and appropriate transcriptional/translational control signals. These methods include *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination. See, for example, the techniques  
15 described in Maniatis *et al.*, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, N.Y. (1989) and Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, N.Y. (1989).

A variety of host-expression vector systems may be utilized to express a  
20 collagen coding sequence. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing a procollagen or collagen coding sequence; yeast or filamentous fungi transformed with recombinant yeast or fungi  
25 expression vectors containing a procollagen or collagen coding sequence; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing sequence encoding the procollagen or collagen of the invention; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid  
30 expression vectors (*e.g.*, Ti plasmid) containing a procollagen or collagen coding sequence; or animal cell systems. The expression elements of these systems vary in their strength and specificities. Depending on the host/vector system utilized, any of a number of suitable transcription and translation elements, including constitutive and  
35 inducible promoters, may be used in the expression vector. For example, when cloning in bacterial systems, inducible promoters such as pL of bacteriophage  $\lambda$ ,

plac, ptrp, ptac (ptrp-lac hybrid promoter) and the like may be used; when cloning in insect cell systems, promoters such as the baculovirus polyhedron promoter may be used; when cloning in plant cell systems, promoters derived from the genome of plant cells (e.g., heat shock promoters; the promoter for the small subunit of RUBISCO; the promoter for the chlorophyll a/b binding protein) or from plant viruses (e.g., the 35S RNA promoter of CaMV; the coat protein promoter of TMV) may be used; when cloning in mammalian cell systems, promoters derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5 K promoter) may be used; when generating cell lines that contain multiple copies of a collagen DNA, SV40-, BPV- and EBV-based vectors may be used with an appropriate selectable marker.

In bacterial systems a number of expression vectors may be advantageously selected depending upon the use intended for the collagen expressed. For example, when large quantities of the collagens of the invention are to be produced for the generation of antibodies, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited to, the *E. coli* expression vector pUR278 (Ruther *et al.*, EMBO J. 2:1791 (1983)), in which the collagen coding sequence may be ligated into the vector in frame with the lac Z coding region so that a hybrid AS-lac Z protein is produced; pIN vectors (Inouye *et al.*, Nucleic Acids Res. 13:3101-3109 (1985); Van Heeke *et al.*, J. Biol. Chem. 264:5503-5509 (1989)); and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S- transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety.

A preferred expression system is a yeast expression system. In yeast, a number of vectors containing constitutive or inducible promoters may be used. For a review see, Current Protocols in Molecular Biology, Vol. 2, Ed. Ausubel *et al.*,

Greene Publish. Assoc. & Wiley Interscience, Ch. 13 (1988); Grant *et al.*,  
Expression and Secretion Vectors for Yeast, in Methods in Enzymology, Ed. Wu &  
Grossman, Acad. Press, N.Y. 153:516-544 (1987); Glover, DNA Cloning, Vol. II,  
IRL Press, Wash., D.C., Ch. 3 (1986); Bitter, Heterologous Gene Expression in  
5 Yeast, in Methods in Enzymology, Eds. Berger & Kimmel, Acad. Press, N.Y.  
152:673-684 (1987); and The Molecular Biology of the Yeast *Saccharomyces*, Eds.  
Strathern *et al.*, Cold Spring Harbor Press, Vols. I and II (1982).

10 A particularly preferred system useful for cloning and expression of the  
collagen proteins of the invention uses host cells from the yeast *Pichia*. Species of  
non-*Saccharomyces* yeast such as *Pichia pastoris* appear to have special advantages in  
producing high yields of recombinant protein in scaled up procedures. Additionally,  
a *Pichia* expression kit is available from Invitrogen Corporation (San Diego, CA).

15 There are a number of methanol responsive genes in methylotrophic yeasts  
such as *Pichia pastoris*, the expression of each being controlled by methanol  
responsive regulatory regions (also referred to as promoters). Any of such methanol  
responsive promoters are suitable for use in the practice of the present invention.  
20 Examples of specific regulatory regions include the promoter for the primary alcohol  
oxidase gene from *Pichia pastoris* AOX1, the promoter for the secondary alcohol  
oxidase gene from *P. pastoris* AXO2, the promoter for the dihydroxyacetone  
synthase gene from *P. pastoris* (DAS), the promoter for the P40 gene from *P.*  
*pastoris*, the promoter for the catalase gene from *P. pastoris*, and the like.

25 Typical expression in *Pichia pastoris* is obtained by the promoter from the  
tightly regulated AOX1 gene. See Ellis *et al.*, Mol. Cell. Biol. 5:1111 (1985) and  
U.S. Patent No. 4,855,231. This promoter can be induced to produce high levels of  
recombinant protein after addition of methanol to the culture. By subsequent  
30 manipulations of the same cells, expression of genes for the collagens of the  
invention described herein is achieved under conditions where the recombinant  
protein is adequately hydroxylated by prolyl 4-hydroxylase and, therefore, can fold  
into a stable helix that is required for the normal biological function of the proteins  
35 in forming fibrils.



Another particularly preferred yeast expression system makes use of the methylotrophic yeast *Hansenula polymorpha*. Growth on methanol results in the induction of key enzymes of the methanol metabolism, namely MOX (methanol oxidase), DAS (dihydroxyacetone synthase) and FMHD (formate dehydrogenase).  
5 These enzymes can constitute up to 30-40% of the total cell protein. The genes encoding MOX, DAS, and FMDH production are controlled by very strong promoters which are induced by growth on methanol and repressed by growth on glucose. Any or all three of these promoters may be used to obtain high level  
10 expression of heterologous genes in *H. polymorpha*. The gene encoding a collagen of the invention is cloned into an expression vector under the control of an inducible *H. polymorpha* promoter. If secretion of the product is desired, a polynucleotide encoding a signal sequence for secretion in yeast, such as the *S. cerevisiae*  
15 prepro-mating factor  $\alpha 1$ , is fused in frame with the coding sequence for the collagen of the invention. The expression vector preferably contains an auxotrophic marker gene, such as URA3 or LEU2, which may be used to complement the deficiency of an auxotrophic host.

20 The expression vector is then used to transform *H. polymorpha* host cells using techniques known to those of skill in the art. An interesting and useful feature of *H. polymorpha* transformation is the spontaneous integration of up to 100 copies of the expression vector into the genome. In most cases, the integrated DNA forms multimers exhibiting a head-to-tail arrangement. The integrated foreign DNA has  
25 been shown to be mitotically stable in several recombinant strains, even under non-selective conditions. This phenomena of high copy integration further adds to the high productivity potential of the system.

Filamentous fungi may also be used to produce the collagens of the instant  
30 invention. Vectors for expressing and/or secreting recombinant proteins in filamentous fungi are well known, and one of skill in the art could use these vectors to express recombinant collagen.

In cases where plant expression vectors are used, the expression of sequences  
35 encoding the collagens of the invention may be driven by any of a number of promoters. For example, viral promoters such as the 35S RNA and 19S RNA

promoters of CaMV (Brisson *et al.*, Nature 310:511-514 (1984), or the coat protein promoter of TMV (Takamatsu *et al.*, EMBO J. 6:307-311 (1987)) may be used; alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi *et al.*, EMBO J. 3:1671-1680 (1984); Broglie *et al.*, Science 224:838-843 (1984); or heat shock promoters, *e.g.*, soybean hsp17.5-E or hsp17.3-B (Gurley *et al.*, Mol. Cell. Biol. 6:559-565 (1986) may be used. These constructs can be introduced into plant cells using Ti plasmids, Ri plasmids, plant virus vectors, direct DNA transformation, microinjection, electroporation, etc. For reviews of such techniques see, for example, Weissbach & Weissbach, Methods for Plant Molecular Biology, Academic Press, NY, Section VIII, pp. 421-463 (1988); and Grierson & Corey, Plant Molecular Biology, 2d Ed., Blackie, London, Ch. 7-9 (1988).

An alternative expression system which could be used to express the collagens of the invention is an insect system. In one such system, *Autographa californica* nuclear polyhidrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. Coding sequence for the collagens of the invention may be cloned into non-essential regions (for example the polyhedron gene) of the virus and placed under control of an AcNPV promoter (for example, the polyhedron promoter). Successful insertion of a collagen coding sequence will result in inactivation of the polyhedron gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedron gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed. (*see, e.g.*, Smith *et al.*, J. Virol. 46:584 (1983); Smith, U.S. Patent No. 4,215,051). Further examples of this expression system may be found in Current Protocols in Molecular Biology, Vol. 2, Ed. Ausubel *et al.*, Greene Publish. Assoc. & Wiley Interscience.

In mammalian host cells, a number of viral based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, coding sequence for the collagens of the invention may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral

genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing collagen in infected hosts. (*See, e.g.*, Logan & Shenk, Proc. Natl. Acad. Sci. USA 81:3655- 3659 (1984)). Alternatively, the vaccinia 7.5 K promoter may be used. (*See, e.g.*, Mackett *et al.*, Proc. Natl. Acad. Sci. USA 79:7415-7419 (1982); Mackett *et al.*, J. Virol. 49:857-864 (1984); Panicali *et al.*, Proc. Natl. Acad. Sci. USA 79:4927-4931 (1982).

Specific initiation signals may also be required for efficient translation of inserted collagen coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where the entire collagen gene, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of a collagen coding sequence is inserted, exogenous translational control signals, including the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the collagen coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (*see* Bittner *et al.*, Methods in Enzymol. 153:516-544 (1987)).

Preferably, the collagens of the invention are expressed as secreted proteins. When the engineered cells used for expression of the proteins are non-human host cells, it is often advantageous to replace the human secretory signal peptide of the collagen protein with an alternative secretory signal peptide which is more efficiently recognized by the host cell's secretory targeting machinery. The appropriate secretory signal sequence is particularly important in obtaining optimal fungal expression of mammalian genes. For example, in methylotrophic yeasts, a DNA sequence encoding the in-reading frame *S. cerevisiae*  $\alpha$ -mating factor pre-pro sequence may be inserted at the amino-terminal end of the coding sequence. The  $\alpha$ MF pre-pro sequence is a leader sequence contained in the  $\alpha$ MF precursor molecule, and includes the lys-arg encoding sequence which is necessary for proteolytic processing and secretion (*see, e.g.*, Brake *et al.*, Proc. Natl. Acad. Sci.

USA 81:4642 (1984)). Other signal sequences for prokaryotic, yeast, fungi, insect or mammalian cells are well known in the art, and one of ordinary skill could easily select a signal sequence appropriate for the host cell of choice.

5 The vectors of this invention may autonomously replicate in the host cell, or may integrate into the host chromosome. Suitable vectors with autonomously replicating sequences ("ars") are well known for a variety of bacteria (*e.g.*, the ars from pBR322 functions in the majority of gram negative bacteria), yeast (the 2 $\mu$  plasmid ars), and various viral replications sequences for both prokaryotes and  
10 eukaryotes (prokaryote:  $\lambda$ , T-even phages, M13, etc; eukaryote: adenovirus, SV40, polyoma, VSV or BPV, vaccina, etc.). Vectors may integrate into the host cell genome when they have a DNA sequence that is homologous to a sequence found in the host cell's genomic DNA.

15 The vectors of the invention also encode a selection gene, also termed a selectable marker, that encodes a product necessary for the host cell to grow and survive under certain conditions. Typical selection genes include genes encoding (1) a protein that confers resistance to an antibiotic or other toxin (*e.g.*, tetracycline, ampicillin, neomycin, methotrexate, etc.), and (2) a protein that complements an  
20 auxotrophic requirement of the host cell, etc. Other examples of selection genes include: the herpes simplex-virus thymidine kinase (Wigler *et al.*, Cell 11:223 (1977)), hypoxanthine-guanine phosphoribosyltransferase (Szybalska *et al.*, Proc. Natl. Acad. Sci. USA 48:2026 (1962)), and adenine phosphoribosyltransferase  
25 (Lowy *et al.*, Cell 22:817 (1980)) genes that can be employed in tk-, hgp<sup>r</sup>t- or ap<sup>r</sup>t-cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler *et al.*, Natl. Acad. Sci. USA 77:3567 (1980); O'Hare *et al.*, Proc. Natl. Acad. Sci. USA 78:1527  
30 (1981)); gpt, which confers resistance to mycophenolic acid (Mulligan *et al.*, Proc. Natl. Acad. Sci. USA 78:2072 (1981)); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin *et al.*, J. Mol. Biol. 150:1 (1981)); and hygro, which confers resistance to hygromycin (Santerre *et al.*, Gene 30:147 (1984)).  
35 Recently, additional selectable genes have been described, namely trpB, which allows cells to utilize indole in place of tryptophan; hisD, which allows cells to utilize

histinol in place of histidine (Hartman *et al.*, Proc. Natl. Acad. Sci. USA **85**:8047 (1988)); and ODC (ornithine decarboxylase) which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine, DFMO (McConlogue L., In: Current Communications in Molecular Biology, Cold Spring Harbor Laboratory, Ed. (1987)).

Further regulatory elements necessary for the expression vectors of the invention include sequences for initiating transcription, *e.g.*, promoters and enhancers. Promoters are untranslated sequences located upstream from the start codon of the structural gene that control the transcription of the nucleic acid under its control. Inducible promoters are promoters that alter their level of transcription initiation in response to a change in culture conditions, *e.g.*, the presence or absence of a nutrient. One of skill in the art would know of a large number of promoters that would be recognized in host cells suitable for the present invention. These promoters are operably linked to the DNA encoding the collagen by removing the promoter from its native gene and placing the collagen encoding DNA 3' of the promoter sequence. Promoters useful in the present invention include, but are not limited to, the following: (prokaryote) (1) the lactose promoter, the alkaline phosphatase promoter, the tryptophan promoter, and hybrid promoters such as the tac promoter, (yeast) (2) the promoter for 3-phosphoglycerate kinase, other glycolytic enzyme promoters (hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, etc.), the promoter for alcohol dehydrogenase, the metallothionein promoter, the maltose promoter, and the galactose promoter, (eukaryotic) (3) virtually all eukaryotic genes have an AT-rich region located approximately 25 to 30 bases upstream from the site where transcription is initiated, examples of suitable eukaryotic promoters include: promoters from the viruses polyoma, fowlpox, adenovirus, bovine papilloma virus, avian sarcoma virus, cytomegalovirus, retroviruses, SV40, and promoters from the target eukaryote including: the glucoamylase promoter from *Aspergillus*, the actin promoter or an immunoglobulin promoter from a mammal, and native collagen promoters. See, *e.g.*, de Boer *et al.*, Proc. Natl. Acad. Sci. USA **80**:21-25 (1983), Hitzeman *et al.*, J. Biol. Chem. **255**:2073 (1980), Fiers *et al.*, Nature **273**:113 (1978), Mulligan and

Berg, Science 209:1422-1427 (1980). Pavlakis *et al.*, Proc. Natl. Acad. Sci. USA 78:7398-7402 (1981), Greenway *et al.*, Gene 18:355-360 (1982), Gray *et al.*, Nature 295:503-508 (1982), Reyes *et al.*, Nature 297:598-601 (1982), Canaani and Berg, Proc. Natl. Acad. Sci. USA 79:5166-5170 (1982), Gorman *et al.*, Proc. Natl. Acad. Sci. USA 79:6777-6781 (1982), Nunberg *et al.*, Mol. and Cell. Biol. 11(4):2306-2315 (1984).

Transcription of the collagen encoding DNA from the promoter is often increased by inserting an enhancer sequence in the vector. Enhancers are cis-acting elements, usually about from 10 to 300 bp, that act to increase the rate of transcription initiation at a promoter. Many enhancers are known for both eukaryotes and prokaryotes, and one of ordinary skill could select an appropriate enhancer for the host cell of interest. See, e.g., Yaniv, Nature 297:17-18 (1982) for eukaryotic enhancers.

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cells lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include, but are not limited to, CHO, VERO, BHK, HeLa, COS, MDCK, 293, WI38, etc. Additionally, host cells may be engineered to express various enzymes to ensure the proper processing of the collagen molecules. For example, the gene for prolyl-4-hydroxylase may be coexpressed with the collagen gene in the host cell.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express the collagens of the invention may be engineered. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with collagen

encoding DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express a desired collagen.

#### **E. Infection, Transformation and Transfection**

Host cells are transfected or preferably infected or transformed with the above-described expression vectors, and cultured in nutrient media appropriate for selecting transductants or transformants containing the collagen encoding vector.

The host cells which contain the coding sequence and which express the biologically active gene product may be identified by at least four general approaches; (a) DNA-DNA or DNA-RNA hybridization; (b) the presence or absence of "marker" gene functions; (c) assessing the level of transcription as measured by the expression of collagen mRNA transcripts in the host cell; and (d) detection of the gene product as measured by immunoassay or by its biological activity.

In the first approach, the presence of the collagen coding sequence inserted in the expression vector can be detected by DNA-DNA or DNA-RNA hybridization using probes comprising nucleotide sequences that are homologous to the collagen coding sequence, respectively, or portions or derivatives thereof.

In the second approach, the recombinant expression vector/host system can be identified and selected based upon the presence or absence of certain "marker" gene functions (*e.g.*, thymidine kinase activity, resistance to antibiotics, resistance to methotrexate, transformation phenotype, occlusion body formation in baculovirus, etc.). For example, if the collagen coding sequence is inserted within a marker gene sequence of the vector, recombinant cells containing collagen coding sequence can be identified by the absence of the marker gene function. Alternatively, a marker gene

can be placed in tandem with the collagen sequence under the control of the same or different promoter used to control the expression of the collagen coding sequence. Expression of the marker in response to induction or selection indicates expression of the collagen coding sequence.

5 In the third approach, transcriptional activity of the collagen coding region can be assessed by hybridization assays. For example, RNA can be isolated and analyzed by Northern blot using a probe homologous to the collagen coding sequence or particular portions thereof. Alternatively, total nucleic acids of the host cell may  
10 be extracted and assayed for hybridization to such probes.

In the fourth approach, the expression of a collagen protein product can be assessed immunologically, for example by Western blots, immunoassays such as radioimmuno-precipitation, enzyme-linked immunoassays and the like.

15

#### F. Purification of Collagens

The expressed collagen of the invention, which is preferably secreted into the culture medium, is purified to homogeneity by chromatography. In one  
20 embodiment, the recombinant collagen protein is purified by size exclusion chromatography. However, other purification techniques known in the art can also be used, including ion exchange chromatography, and reverse-phase chromatography. See, e.g., Maniatis *et al.*, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, N.Y. (1989), Ausubel *et al.*, Current Protocols in Molecular  
25 Biology, Greene Publishing Associates and Wiley Interscience, N.Y. (1989), and Scopes, Protein Purification: Principles and Practice, Springer-Verlag New York, Inc., NY (1994).

The present invention is further illustrated by the following examples, which  
30 are not intended to be limited in any way.

### VI. **EXAMPLES**

#### A. **Example 1: Synthesis of Human Type II Procollagen**

35 A recombinant COL1A1 gene construct employed in the present invention comprised a fragment of the 5'- end of COL1A1 having a promotor, exon



1 and intron 1 fused to exons 3 through 54 of a COL2A1 gene. The hybrid construct was transfected into HT-1080 cells. These cells were co-transfected with a neomycin-resistance gene and grown in the presence of the neomycin analog G418.

5 The hybrid construct was used to generate transfected cells.

A series of clones were obtained that synthesized mRNA for human type II procollagen. To analyze the synthesized proteins, the cells were incubated with [ $^{14}\text{C}$ ] proline so that the medium proteins could be analyzed by autoradiography (storage phosphor film analyzer).

10 As set forth at Figure 1, lane 1 shows that the unpurified medium proteins are comprised of three major polypeptide chains. Specifically, the medium proteins contained the expected type II procollagen comprised of  $\text{pro}\alpha 1(\text{II})$  chains together with  $\text{pro}\alpha 1(\text{IV})$  and  $\text{pro}\alpha 2(\text{IV})$  chains of type IV collagen normally synthesized by  
15 the cells. The upper two are  $\text{pro}\alpha 1(\text{IV})$  and  $\text{pro}\alpha 2(\text{IV})$  chains of type IV collagen that are synthesized by cells not transfected by the construct. The third band is the  $\text{pro}\alpha 1(\text{II})$  chains of human type II procollagen synthesized from the construct. Lanes 2 and 3 are the same medium protein after chromatography of the medium on an ion  
20 exchange column. As indicated in Lanes 2 and 3, the type II procollagen was readily purified by a single step of ion exchange chromatography.

The type II procollagen secreted into the medium was correctly folded by a protease-thermal stability test. As evidenced at Figure 2, the medium proteins were digested at the temperatures indicated with a high concentration of trypsin and  
25 chymotrypsin under conditions in which correctly folded triple-helical procollagen or collagen resists digestion but unfolded or incorrectly folded procollagen of collagen is digested to small fragments. The products of the digestion were then analyzed by polyacrylamide gel electrophoresis in SDS and fluorography. The results show that  
30 the type II procollagen resisted digestion up to 43°C, the normal temperature at which type II procollagen unfolds. Therefore, the type II procollagen is correctly folded and can be used to generate collagen fibrils.

35

**B. Example 2: Synthesis of Human Type I Procollagen**

As a second example, HT-1080 cells were co-transfected with a COL1A1 gene and a COL1A2 gene. Both genes consisted of a cytomegalic virus promoter linked to a full-length cDNA. The COL1A2 gene construct but not the COL1A1 gene construct contained a neomycin-resistance gene. The cells were selected for expression of the COL1A2- neomycin resistance gene construct by growth in the presence of the neomycin-analog G418. The medium was then examined for expression of the COL1A1 with a specific polyclonal antibody for human pro $\alpha$ 1(I) chains.

More specifically, the COL1A2 was linked to an active neomycin-resistance gene but the COL1A1 was not. The cells were screened for expression of the COL1A2-neomycin resistance gene construct with the neomycin analog G418. The medium was analyzed for expression of the COL1A1 by Western blotting with a polyclonal antibody specific for the human pro $\alpha$ 1(I) chain. As set forth in Figure 3, lane 1 indicates that the medium proteins contained pro $\alpha$ (I) chains ( $\alpha$ 1(I) and  $\alpha$ 2(I)). Lane 2 is an authentic standard of type I procollagen containing pro $\alpha$ 1(I) and pro $\alpha$ 2(I) chains and partially processed pC $\alpha$ 1(I) chains. The results demonstrate that the cells synthesized human type procollagen that contained pro $\alpha$ 1(I) chains, presumably in the form of the normal heterotrimer with the composition two pro $\alpha$ 1(I) chains and one pro $\alpha$ 2(I) chain.

These results demonstrated that the cells synthesized human type I procollagen that was probably comprised of the normal heterotrimeric structure of two pro $\alpha$ 1(I) chains and one pro $\alpha$ 2(I) chain.

TABLE I presents a summary of some of the DNA constructs containing human procollagen genes. The constructs were assembled from discrete fragments of the genes or cDNAs from the genes together with appropriate promoter fragments.

**TABLE I**

<u>Constructs</u>	<u>5'end</u>	<u>Central Region</u>	<u>3'end</u>	<u>Protein product</u>
A	Promoter (2.5 kb) + exon 1 + intron 1 from COL1A1	Exons 3 to 54 from COL2A 1	3.5 kb SphI/SphI fragment from 3'end of COL2A1	Human type II procollagen, [pro $\alpha$ 1(II)] <sup>3</sup>
B	Promoter (2.5 kb) of COL1A1	Exons 1 to 54 from COL2A 1	3.5 kb SphI/SphI fragment from 3'end of COL2A1	Human type II procollagen [pro $\alpha$ 1(II)]
C	Promoter (2.5 kb) + exon 1 + intron 1 + half of exon 2 from COL1A1	cDNA for COL1A 1 except for first 1 ½ exons	0.5 kb fragment from COL1A1	Human type I procollagen, [pro $\alpha$ 1(I)] <sub>3</sub>
D	Cytomegalic virus promoter	cDNA from COL1A 1		Human type I procollagen, [pro $\alpha$ 1(I)] <sub>3</sub>
E	Cytomegalic virus promoter	cDNA from COL1A 2		Human type I [pro $\alpha$ 1(I)] <sub>2</sub> pro $\alpha$ 2(I)] when expressed with construct C or D

**C. Example 3: Cell Transfections**

For cell transfection experiments, a cosmid plasmid clone containing  
 35 the gene construct was cleaved with a restriction endonuclease to release the  
 construct from the vector. A plasmid vector comprising a neomycin resistance gene,

(Law *et al.*, Mol. Cell. Biol. 3:2110-2115 (1983)) was linearized by cleavage with BamHI. The two samples were mixed in a ratio of approximately 10:1 gene construct to neomycin resistant gene, and the mixture was then used for cotransfection of HT-1080 cells by calcium phosphate coprecipitation (Sambrook *et al.*, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory Press, 2d Edition (1989)). DNA in the calcium phosphate solution was layered onto cultured cells without 10 $\mu$ g of chimeric gene construct per 100 ml plate of preconfluent cells. Cells were incubated in DMEM containing 10% newborn calf serum for 10 hours. The samples were subjected to glycerol shock by adding a 15% glycerol solution for 3 minutes. The cells were then transferred to DMEM medium containing newborn calf serum for 24 hours and then to the same medium containing 450  $\mu$ g/ml of G418. Incubation in the medium containing G418 was continued for about 4 weeks with a change of medium every third day. G418-resistant cells were either pooled or separate clones obtained by isolating foci with a plastic cylinder and subcultured.

#### 20 D. Example 4: Western Blotting

For assay of expression of the COL2A1 gene, polyclonal antibodies were prepared in rabbits using a 23-residue synthetic peptide that had an amino acid sequence found in the COOH-terminal telopeptide of type II collagen. See generally, Cheah *et al.*, Proc. Natl. Acad. Sci. USA 82:2555-2559 (1985). The antibody did not react by Western blot analysis with pro $\alpha$  chains of human type I procollagen or collagen, human type II procollagen or collagen, or murine type I procollagen. For assay of expression of the COL1A1 genes, polyclonal antibodies that reacted with the COOH-terminal polypeptide of the pro $\alpha$ (I) chain were employed. See generally, Olsen *et al.*, J. Biol. Chem. 266:1117-1121 (1991).

Culture medium from pooled clones or individual clones was removed and separately precipitated by the addition of solid ammonium sulfate to 30% saturation and precipitates were collected by centrifugation at 14,000 x g and then dialyzed against a buffer containing 0.15 M NaCl, 0.5 mM EDTA, 0.5 mM N-ethylmaleimide, 0.1 mM and p-aminobenzamidine, and 50 mM Tris-HCl (pH 7.4

at 4°C). Aliquots of the samples were heated to 10°C for 5 minutes in 1% SDS, 50 mM DTT and 10% (v/v) glycerol, and separated by electrophoresis on 6% polyacrylamide gels using a mini-gel apparatus (Holford SE250, Holford Scientific) run at 125 V for 90 minutes. Separated proteins were electroblotted from the polyacrylamide gel at 40 V for 90 minutes onto a supported nitrocellulose membrane (Schleicher and Schuell). The transferred proteins were reacted for 30 minutes with the polyclonal antibodies at a 1:500 (v/v) dilution. Proteins reacting with the antibodies were detected with a secondary anti-rabbit IgG antibody coupled to alkaline phosphatase (Promega Biotech) for 30 minutes. Alkaline phosphatase was visualized with NBT/BCIP (Promega Biotech) as directed by the manufacturer.

E. Example 5: *In vitro* Analysis Of Recombinant Collagen.

1. Assembly Of Recombinant Collagen: Protease Digestion.

To demonstrate that the procollagens synthesized and secreted in the medium by the transfected cells were correctly folded, the medium proteins were digested with high concentrations of proteases under conditions in which only correctly folded procollagens and collagens resist digestion. For digestion with a combination of trypsin and chymotrypsin, the cell layer from a 25 cm flask was scraped into 0.5 ml of modified Krebs II medium containing 10 mM EDTA and 0.1% Nonidet P-40 (Sigma). The cells were vigorously agitated in a Vortex mixer for 1 minute and immediately cooled to 4°C. The supernatant was transferred to new tubes. The sample was preincubated at the temperature indicated for 10 minutes and the digestion was carried out at the same temperature for 2 minutes. For the digestion, a 0.1 volume of the modified Krebs II medium containing 1 mg/ml trypsin and 2.5 mg/ml  $\alpha$ -chymotrypsin (Boehringer Mannheim) was added. The digestion was stopped by adding a 0.1 volume of 5 mg/ml soybean trypsin inhibitor (Sigma). For analysis of the digestion products, the sample was rapidly immersed in boiling water for 2 minutes with the concomitant addition of a 0.2 volume of 5 x electrophoresis sample buffer that consisted of 10% SDS, 50% glycerol, and 0.012% bromphenol blue in 0.625 M Tris-HCl buffer (pH 6.8). Samples were applied to SDS gels with prior reduction by incubating for 3 minutes in boiling water after the

addition of 2% 2-mercaptoethanol. Electrophoresis was performed using the discontinuous system of Laemli, Nature 227:680-685 (1979), with minor modifications described by de Wet *et al.*, J. Biol. Chem. 258:7721-7728 (1983).

5

## 2. Double Immunostaining Of Sf9 Cells.

Sf9 cells were grown on glass slides and fixed in 100% ethanol at -20°C. Alternatively, cells in monolayer were detached, washed twice with a solution of 0.15 M NaCl and 0.02 M phosphate, pH 7.4 (washing solution),  
10 suspended in cold ethanol and spread on silanated (Maples, J.A., (1985), Am. J. Clin. Pathol. 83:356- 363) glass slides. Cells were incubated with 1% bovine serum albumin in 0.15 M NaCl and 0.02 M phosphate, pH 7.4, for 15 min followed by incubation for 30 min in a 1:50 dilution of a mouse monoclonal antibody to the  $\beta$   
15 subunit (5B5, Dako) and a rabbit polyclonal antibody to the  $\alpha$  subunit of human prolyl 4-hydroxylase in the above bovine serum albumin-containing solution. Cells were washed with the washing solution 4 times for 20 min and incubated in a 1:10 dilution of a sheep anti-mouse Ig-rhodamine F(ab)2 fragment (Boehringer Mannheim)  
20 and a sheep anti- rabbit IgG fluorescein F(ab)2 fragment (Boehringer Mannheim) in the bovine serum albumin-containing solution for 30 min, washed with the washing solution, rinsed with distilled water and mounted using Glycergel (Dako). The samples were photographed using a Leitz Aristoplan microscope equipped with ep-illuminator and filters for fluorescein isothiocyanate and tetramethyl rhodamine B  
25 isothiocyanate fluorescence.

To study the efficiency of a multiple baculovirus infection, immunocytochemical staining of insect cells was used. Sf9 cells were coinfectd with two recombinant viruses coding for the ( and ( subunits of prolyl 4-hydroxylase  
30 and immunostained with antibodies to these two subunits (Fig. 3). When the analysis was performed 48 h after infection, 87% of all cells were found to express at least one of the two types of subunit, 90% of cells expressing one type of subunit also expressing the other type.

35

### 3. *Prolyl 4-Hydroxylase Activity Assay.*

The 0.2% Triton X-100 extracts of cell homogenates were analyzed for prolyl 4-hydroxylase activity with an assay based on the hydroxylation-coupled decarboxylation of 2-oxo [1-14C] glutarate (Kivirikko *et al.*, Methods Enzymol. 82:245-304 (1982)). As reported previously (Veijola *et al.*, J. Biol. Chem. 269:26746-26753 (1994)), a significant level of prolyl 4-hydroxylase activity was found in both Sf9 and High Five cells, the activity in High Five cells being distinctly higher than that in Sf9 cells (TABLE I). Infection of the cells with a virus coding for the pro $\alpha$ 1 (III) chains had only minor effects on this activity, whereas the activity in cells infected with the virus coding for the pro $\alpha$ 1 (III) chain together with viruses coding for the two types of subunit of human prolyl 4-hydroxylase was markedly higher (TABLE I).

### 4. *Assay For Measuring Collagen.*

The amount of the purified type III collagen was determined by using the Sircol collagen assay (Biocolor). Amino acid analysis of the purified type III collagen was performed in an Applied Biosystems 421 Amino Acid Analyzer.

### F. *Example 6: Specifically Engineered Procollagens and Collagens*

As indicated in Figure 4, a hybrid gene consisting of some genomic DNA and some cDNA for the pro $\alpha$ 1(I) chain of human type I procollagen was the starting material. The DNA sequence of the hybrid gene was analyzed and the codons for amino acids that formed the junctions between the repeating D-periods were modified in ways that did not change the amino acids encoded but did create unique sites for cleavage of the hybrid gene by restriction endonucleases.

#### 1. *Recombinant Procollagen Or Collagen*

The D3-period of pro $\alpha$ 1(I) is excised using SrfI and NaeI restriction nucleases. The bases coding for the amino acids found in the collagenase recognition site present in the D3 period are modified so that they code for a different amino acid sequence. The cassette is amplified and reinserted in the gene.

Expression of the gene in an appropriate host cell will result in type I collagen which cannot be cleaved by collagenase.

5                   2.     ***Procollagen Or Collagen Deletion Mutants***

A D2 period cassette (of the pro $\alpha$ 1(I) chain) is excised from the gene described above by digestion with SmaI. The gene is reassembled to provide a gene having a specific 5 in-frame deletion of the codons for the D-2 period.

10                   3.     ***Procollagen Or Collagen Addition Mutants***

Multiple copies of one or more D-cassettes may be inserted at the engineered sites to provide multiple copies of desired regions of procollagen or collagen.

15                   G.     **Example 7: Expression Of Human Prolyl 4-Hydroxylase In A Recombinant DNA System**

To obtain expression of the two genes for prolyl 4-hydroxylase in insect cells, the following procedures were carried out. The baculovirus transfer  
20 vector pVL $\alpha$ 58 was constructed by digesting a pBluescript (Stratagene) vector containing in the SmaI site the full-length cDNA for the  $\alpha$  subunit of human prolyl 4- hydroxylase, P $\alpha$ -58 (Helaakoski *et al.*, Proc. Natl. Acad. Sci. USA 86, 4392-4396 (1989)), with PstI and BamHI, the cleavage sites which closely flank the  
25 SmaI site. The resulting PstI-PstI and PstI-BamHI fragments containing 61 bp of the 5' untranslated sequence, the whole coding region, and 551 bp of the 3' untranslated sequence were cloned to the PstI-BamHI site for the baculovirus transfer vector pVL1392 (Luckow *et al.*, Virology 170:31-39 (1989)). The baculovirus transfer  
30 vector pVL $\alpha$ 59 was similarly constructed from pVL1392 and another cDNA clone, P $\alpha$ -59 (Helaakoski *et al.*, *supra*), encoding the  $\alpha$  subunit of human prolyl 4-hydroxylase. The cDNA clones P $\alpha$ -58 and P $\alpha$ -59 differ by a stretch of 64 bp.

The pVL $\beta$  vector was constructed by ligation of an EcoRI-BamHI fragment  
35 of a full-length cDNA for the  $\beta$  subunit of human prolyl 4-hydroxylase, S-138 (Pihlajaniemi *et al.*, EMBO J. 6:643-649 (1987)) containing 44 bp of the 5' untranslated sequence, the whole coding region, and 207 bp of the 3' untranslated



sequence to EcoRI/BamHI-digested pVL1392. Recombinant baculovirus transfer vectors were cotransfected into Sf9 cells (Summers *et al.*, Tex. Agric. Exp. St. Bull. 1555:1-56 (1987)) with wild-type Autographa californica nuclear polyhedrosis virus (AcNPV) DNA by calcium phosphate transfection. The resultant viral pool in the supernatant of the transfected cells was collected 4 days later and used for plaque assay. Recombinant occlusion-negative plaques were subjected to three rounds of plaque purification to generate recombinant viruses totally free of contaminating wild-type virus. The screening procedure and isolation of the recombinant viruses essentially followed by the method of Summers and Smith, *supra*. The resulting recombinant viruses from pVL $\alpha$ 58, pVL $\alpha$ 59, and pVL $\beta$  were designated as the  $\alpha$ 58 virus,  $\alpha$ 59 virus and  $\beta$  virus, respectively.

Sf9 cells were cultured in TNM-FH medium (Sigma) supplemented with 10% fetal bovine serum at 27°C either as monolayers or in suspension in spinner flasks (Techne). To produce recombinant proteins, Sf9 cells seeded at a density of 106 cells per ml were injected at a multiplicity of 5-10 with recombinant viruses when the  $\alpha$ 58,  $\alpha$ 59, or  $\beta$  virus was used alone. The  $\alpha$  and  $\beta$  viruses were used for infection in ratios of 1:10-10:1 when producing the prolyl 4-hydroxylase tetramer. The cells were harvested 72 hours after infection, homogenized in 0.01 M Tris, pH 7.8/0.1 M NaCl/0.1 M glycine/10 $\mu$ M dithiothreitol/0.1% Triton X-100, and centrifuged. The resulting supernatants were analyzed by SDS/10% PAGE or nondenaturing 7.5% PAGE and assayed for enzyme activities. The cell pellets were further solubilized in 1% SDS and analyzed by SDS/10% PAGE. The cell medium at 24-96 hours postinfection was also analyzed by SDS/10% PAGE to identify any secretion of the resultant proteins into the medium. The cells in these experiments were grown in TNM-FH medium without serum.

When the time course of protein expression was examined, Sf9 cells infected with recombinant viruses were labeled with [ $^{35}$ S]methionine (10  $\mu$ Ci/ $\mu$ l; Amersham; 1 Ci=37CBq) for 2 hours at various time points between 24 and 50 hours after infection and collected for analysis by SDS/10% PAGE. To determine the maximal accumulation of recombinant protein, cells were harvested at various times from 24 to 96 hours after infection and analyzed on by SDS/10% PAGE. Both the 0.1%

Triton X-100- and 1% SDS-soluble fractions of the cells were analyzed. Prolyl 4-hydroxylase activity was assayed by a method based on the decarboxylation of 2-oxo[1-14C]glutarate (Kivirikko *et al.*, Methods in Enzymology 82:245-304 (1982)).

5 The  $K_m$  values were determined by varying the concentrations of one substrate in the presence of fixed concentration of the second, while the concentrations of the other substrates were held constant (Myllyla *et al.*, Eur. J. Biochem. 80:349-357 (1977)).

Protein disulfide-isomerase activity of the  $\beta$  subunit was measured by glutathione: insulin transhydrogenase assay (Carmichael *et al.*, J. Biol. Chem. 252:7163-7167

10 (1977)). Western blot analysis was performed using a monoclonal antibody, 5B5, to the  $\beta$  subunit of human prolyl 4-hydroxylase (Hoyhtya *et al.*, Eur. J. Biochem. 141:477-482 (1984)). Prolyl 4-hydroxylase was purified by a procedure consisting of poly (L-proline) affinity chromatography, DEAE-cellulose chromatography, and gel

15 filtration (Kivirikko *et al.*, Methods in Enzymology 144:96-114 (1987)).

Figure 5 presents analysis of the prolyl 4-hydroxylase synthesized by the insect cells after purification of the protein by affinity-column chromatography. When examined by polyacrylamide gel electrophoresis in a nondenaturing gel, the

20 recombinant enzyme co-migrated with the tetrameric and active form of the normal enzyme purified from chick embryos. After the purified recombinant enzyme was reduced, the (- and (- subunits were detected. As set forth in Figure 5, lanes 1-3 are protein separated under non-denaturing conditions and showing tetramers of the two

25 kinds of subunits. Lanes 4-6 are the same samples separated under denaturing conditions so that the two subunits appear as separate bands.

TABLE II presented data on the enzymic activity of the recombinant enzyme. The  $K_m$  values were determined by varying the concentration of one substrate in the presence of fixed concentrations of the second while the concentration of the other

30 substrates were held constant.

**TABLE II**

Substrate	Km value, $\mu$ M		
	$\alpha$ 58282	$\alpha$ 59282	Chick enzyme
Fe+2	4	4	4
2-oxoglutarate	22	25	22
ascorbate	330	330	300
(Pro-Pro-Gly)10	18	18	15-20

As indicated, the Michales-Mento (Km) values for the recombinant enzyme were essentially the same as for the authentic normal enzyme from chick embryos.

Since the transfected insect cells synthesize large amounts of active prolyl 4-hydroxylase, they are appropriate cells to transfect with genes of the present invention coding for procollagens and collagens so as to obtain synthesis of large amounts of the procollagens and collagens. Transfection of the cells with genes of the present invention is performed as described in Example 3.

#### H. Example 8: Expression of Recombinant Collagen Genes in *Saccharomyces cerevisiae* Yeast Expressing Recombinant Genes for Prolyl 4-Hydroxylase

The yeast *Saccharomyces cerevisiae* can be used with any of a large number of expression vectors. One of the most commonly employed expression vectors is the multi-copy  $2\mu$  plasmid that contains sequences for propagation both in yeast and *E. coli*, a yeast promoter and terminator for efficient transmission of the foreign gene. Typical examples of such vectors based on  $2\mu$  plasmids are pWYG4 that has the  $2\mu$  ORI-STB elements, the GALI promoter, and the  $2\mu$  D gene terminator. In this vector an NcoI cloning site is used insert the gene for either the ( or ( subunit of prolyl 4-hydroxylase, and provide the ATG start codon for either the  $\alpha$  or  $\beta$  subunit. As another example, the expression vector can be pWYG7L that has intact  $2\mu$  ORI, STB, REP1 and REP2, the GAL7 promoter, and uses the FLP terminator. In this vector, the gene for either the ( or ( subunit of prolyl 4-hydroxylase is inserted in the polylinker with its 5' ends at a BamHI or NcoI site.

The vector containing the prolyl 4-hydroxylase gene is transformed into *S. cerevisiae* either after removal of the cell wall to produce spheroplasts that take up DNA on treatment with calcium and polyethylene glycol or by treatment of intact cells with lithium ions. Alternatively, DNA can be introduced by electroporation.

Transformants can be selected by using host yeast cells that are auxotrophic for leucine, tryptophane, uracil or histidine together with selectable marker genes such as LEU2, TRO1, URA3, HIS3 or LEU2-D. Expression of the prolyl 4-hydroxylase genes driven by the galactose promoters can be induced by growing the culture on a non-repressing, non-inducing sugar so that very rapid induction follows addition of galactose; by growing the culture in glucose medium and then removing the glucose by centrifugation and washing the cells before resuspension in galactose medium; and by growing the cells in medium containing both glucose and galactose so that the glucose is preferentially metabolized before galactose-induction can occur. Further manipulations of the transformed cells are performed as described above to incorporate genes for both subunits of prolyl 4-hydroxylase and desired collagen or procollagen genes into the cells to achieve expression of collagen and procollagen that is adequately hydroxylated by prolyl 4-hydroxylase to fold into a stable triple helical conformation and therefore accompanied by the requisite folding associated with normal biological function.

**I. Example 9 Expression of Recombinant Collagen Genes in *Pichia pastoris* Yeast Expressing Recombinant Genes for Prolyl 4-Hydroxylase**

Expression of the genes for prolyl 4-hydroxylase and procollagens or collagens can also be in non-*Saccharomyces* yeast such as *Pichia pastoris* that appear to have special advantages in producing high yields of recombinant protein in scaled-up procedures. Typical expression in the methylotroph *P. pastoris* is obtained by the promoter from the tightly regulated AOX1 gene that encodes for alcohol oxidase and can be induced to give high levels of recombinant protein driven by the promoter after addition of methanol to the cultures. Since *P. Pastoris* has no native plasmids, the yeast is employed with expression vectors designed for chromosomal integration and genes such as HIS4 are used for selection. By subsequent

manipulations of the same cells, expression of genes for procollagens and collagens described herein is achieved under conditions where the recombinant protein is adequately hydroxylated by prolyl 4-hydroxylase and, therefore, can fold into a stable helix that is required for the normal biological function of the proteins in forming fibrils.

The following vectors have been constructed according to the disclosures set forth herein:

1. Human collagen type III without its own signal sequence.

The 3' end of the collagen DNA was synthesized from 4195 bp downstream (*EcoRI* site) of the translation initiation codon to stop codon (4401 bp) of the translation by PCR (see Ala-Kokko *et al.*, 1989 *Biochem J.*, 260:509-516 accession number X144207) using pBluescript-SM38. *NotI* and *XbaI* sites were created in the 3' end of the fragment. pBluescript-SM38 was digested with *EcoRI-XbaI* and the large fragment (approximately 7.2 kb) was isolated. This large *EcoRI-XbaI* fragment (approximately 7.2 kb) and the 3' collagen PCR *EcoRI-XbaI* fragment were ligated with T4 ligase to give the plasmid pBluescript-SM38B.

The 5' end of the collagen DNA was synthesized from 73bp downstream of the translation initiation codon to 176 bp (*BamHI* site) by PCR (for sequences, see Ala-Kokko *et al.*, 1989 *Biochem J.*, 260, 509-516). *ClaI* and *NotI* sites were created in the 5' end of the fragment. pbluescript-SM38/B was digested by *ClaI* and *BamHI* and the large fragment (approximately 6.3 kb) and the 862 bp *BamHI-BamHI* collagen fragment were isolated. The 5' PCR collagen fragment *BamHI-ClaI*, large fragment *BamHI-ClaI* (approximately 6.3 kb) and 862 bp *BamHI-BamHI* fragments were ligated with T4 ligase (triple ligation), to give plasmid pBluescript-SM38/11.

pBluescript-SM38 11 was digested by *NotI*. Then *NotI-NotI* collagen fragment (73 bp-4401 bp) was cloned in frame with  $\alpha$ -factor signal sequence into the pPIC9 (Invitrogen), to give plasmid pPIC9Col III(clone7).

2. Human collagen type III with its own signal sequence

The 3' end of the collage was created from 4195 bp downstream (EcoRI site) of the translation initiation codon to the stop codon of the translation by PCR using pBluescript-SM38. An XbaI site was created in the 3' end of the fragment. The ensuing pBluescript-C3A1 plasmid was digested by EcoRI and XbaI, large fragment (approximately 7.2 kb) was isolated and the large fragment (approximately 7.2 kb) and the 3' PCR collagen fragment were ligated with T4, to give plasmid pBluescript-C3A1/10. A Bg/II site was created 16 bp upstream of the translation initiation codon (Lamberg *et al.*, 1996). Bg/II-XbaI collagen fragment (-16 bp - 4401 bp) of pBluescript-C3A1/10 was then ligated into the EcoRI site of the pHIL-D2 (Invitrogen) to give plasmid pHII- D2/collIII.

3. Human prolyl 4-hydroxylase

A vector, pYM25, which contains ARG4 gene of *Saccharomyces Cerevisiae*, was digested by HpaI. The HpaI fragment of ARG4 gene was inserted into the EcoRV sites of pAO815 (Invitrogen) to create a vector pARG815, which contains the ARG4 gene instead of the HIS4 gene.

The  $\beta$ -subunit cDNA of prolyl 4-hydroxylase (Vuori *et al.*, 1992 PNAS., 89, 7467-7470) was synthesized from the translation initiation codon to the stop codon by PCR. EcoRI restriction sites were created in 5' and 3' ends. The C-terminal ER retention peptide sequence-KDEL-was modified to yeast ER retention signal-HDEL-(CAC GAT GAA CTG) by PCR. The EcoRI-EcoRI \_-fragment was inserted into the pBluescript SK, to give plasmid pBluescript SK $\beta$ /20. Then this EcoRI \_-fragment was inserted into the EcoRI site of PAO815 (Invitrogen) to create single expression cassette vector.

The 5' end of the  $\alpha$ -subunit was synthesized from the translation initiation codon to 689 bp downstream (HindIII site) by PCR. HindIII and SmaI sites were created in 5' end of the fragment. Plasmid pA-59 (see Vuori *et al*) was digested by HindIII and the large fragment (approximately 4.9 kb) was isolated. The large fragment (approximately 4.9 kb) and the 5' PCR fragment were ligated with T4 ligase, to give plasmid pA-59/15.

The 3' end was synthesized from 1373 bp (*Pst*I site) downstream of the translation initiation to the translation stop codon. *Sma*I and *Bam*HI sites were created in 3' end of the fragment. Plasmid pA-59/15 was digested with *Pst*I and *Bam*HI, the large fragment (approximately 3.9 kb) was isolated, and the large fragment and 3' PCR fragment were ligated with T4 ligase, to give plasmid pA-59/3. Plasmid pA-59 was digested with *Sma*I and the *Sma*I-*Sma*I  $\alpha$ -subunit fragment (1 bp-1605 bp) was ligated into *Eco*RI site of the pARG815.

The  $\beta$  single cassette vector was digested by *Bg*/II-*Bam*HI to excise the expression cassette and the expression cassette was reinserted into one for the *Bam*HI site of pARG815 $\alpha$  expression vector. Thereby the vector contains two expression cassettes: one for the  $\alpha$ -subunit and one for the  $\beta$ -subunit.

The  $\beta$ -subunit without its signal sequence was synthesized by PCR from 52 bp downstream of the translation initiation codon to the translation stop codon. *Eco*RI restriction sites were created in 5' and 3' ends. This PCR fragment was cloned into the *Eco*RI site of pSP72 (Promega).

#### J. Example 10: Expression of Recombinant Collagen Genes in Insect Cells Expressing Recombinant Genes for Prolyl 4-Hydroxylase

##### 1. Construction of Recombinant Vectors Containing Collagen Genes.

pVLC1A1: The baculovirus transfer vector was constructed using the eukaryotic expression vector CMV- COL1A1 (Geddis *et al.*, Matrix 13:399-405 (1993)) and the polyhedrin-based baculovirus transfer vector pVL 1392 (Luckow *et al.*, Virology 170:31-39 (1989)). CMV-COL1A1 contains the sequences coding for the full length cDNA sequence of the  $\alpha$ 1 chain of the human procollagen I (COL1A1).

Digestion of CMV-COL1A1 with *Xba*I generates the full length cDNA for COL1A1 including six bp 5' untranslated, and 222 bp 3' untranslated, and this fragment is cloned into the *Xba*I site of pVL1392 to give the plasmid pVLC1A1.

pVLC1A2: The baculovirus transfer vector was constructed using the vector pUC-HP2010 (Kuivaniemi *et al.*, Biochem. J. 252:633-640 (1988)) and the polyhedrin-based baculovirus transfer vector pVL 1392 (Luckow *et al.*, Virology

170:31-39 (1989)). pUC-HP2010 contains the sequences coding for the full length cDNA sequence of the  $\alpha 2$  chain of the human procollagen I (COL1A2) in the *SphI* site of pUC19.

5 pUC-HP2010 is digested with *SphI*, the GTAC overhang is removed with T4 DNA Polymerase, and the blunt ended fragment is cloned into the *EcoRV* site of pSP72 (Promega). A *BglII* site is made six bp upstream of the translation initiation site by PCR, to give the plasmid pSP72-C1A2T. The full length cDNA for COL1A2 is generated by cutting pSP72-C1A2T with *BglII-BamHI*. The *BglII-BamHI* fragment 10 from pSP72-C1A2T has the full length COL1A2 sequence plus six bp 5' untranslated, and 278 bp 3' untranslated, and this fragment is cloned into the *BglII-BamHI* sites of pVL1392 to give pVLC1A2.

pVLC3A1: A *BglII* site was created 16 bp upstream of the translation 15 initiation codon to a full-length cDNA including 92 bp 5' untranslated region and 715 bp 3' untranslated region for the pro $\alpha 1$  chain of human type III procollagen in the plasmid pBS-SM38 (derived from sequences presented in Ala-Kokko *et al.* Biochem. J. 260: 509-516 (1989), and GenBank accession number X14420) by 20 PCR, to give the plasmid pBS-C3A1. pBS-C3A1 was digested with *BglII* and *XbaI* restriction enzymes and the *BglII/XbaI* fragment containing the full-length cDNA of pro $\alpha 1$  chain of human type III procollagen including 16 bp 5' untranslated region, and 715 bp 3' untranslated region, was then ligated to pVL1392 (Luckow *et al.* Virology 170:31-39 (1989)) to give the plasmid pVLC3A1.

25 pVLC3A15'UT/C2A1: The baculovirus transfer vector was constructed using the sequences presented in Baldwin *et al.*, Biochem. J. 262:521-528 (1989) resulting in the vector pGEMC2A1 and the polyhedrin-based baculovirus transfer vector pVL 1392 (Luckow *et al.*, Virology 170:31-39 (1989)). pGEMC2A1 contains the 30 sequences coding for exon I from type I collagen, and type II collagen starts from exon 2B.

pGEMC2A1 is digested with *XbaI-DraI* to generate a fragment with the full length cDNA fusion, and six bp 5' untranslated region and 396 bp 3' untranslated 35 region, and this fragment is cloned into the *XbaI-SmaI* sites of pVL1392 to give the plasmid pVLC1A1/C2A1. The 5' untranslated region was then changed to



GATCTGATATT by cloning an oligonucleotide into the *Bgl*III-*Xba*I sites of the COL II vector.

pVLC3A1NP/C2A1: pGEMC2A1 is digested with *Xba*I-*Bam*HI and the full length cDNA fusion is cloned into the *Xba*I-*Bam*HI sites of pBS(SK-) to give the plasmid pBSC1A1/C2A1. pBSC1A1/C2A1 is digested with *Bgl*III-*Nar*I to generate a full length cDNA without the N-propeptide, the N-propeptide with 16 bp 5' untranslated from type III collagen was synthesized by PCR using the plasmid pBS-C3A1 as a template. The oligonucleotides used to synthesize the type-III N-propeptide were as follows: 5' oligo (5'-TACTCTAGACTCAGATCTGATATT-3') and 3' oligo (5'-GGGAGAATAGTTCTGAGGACCAGT-3'). The 35 bp fragment of telopeptide from type II collagen was synthesized by oligonucleotides (chemical synthesis). The following oligonucleotides were used 5'-CAGATGGCTGGAGGATTTGATGAAAAGG CT GGTGG-3'; and 5'-CGCCACCAGCCTTTTCATCAAATCCTCCAGCCATCTG-3'. These fragments were ligated into pBSC1A1/C2A1 digested with *Bgl*III-*Nar*I. This hybrid full length cDNA was excised with *Bgl*III-*Dra*I and cloned into the *Bgl*III-*Not*I (the *Not*I site is blunt ended by filling in the overhangs with klenow and dNTPs) sites of pVL1392 to give the plasmid pVLC3A1NP/C2A1.

pVLC4A1: The baculovirus transfer vector was constructed using the vector  $\alpha$ 1CMVC which was constructed by R. Niecht K\*ln (based on the sequence published by Brazel *et al.*, Eur. J. Biochem. 168:529-536 (1987), and Soininen *et al.*, FEBS Lett. 225:188-194 (1987)) and the polyhedrin-based baculovirus transfer vector pVL 1392 (Luckow *et al.*, Virology 170:31-39 (1989)).

$\alpha$ 1CMVC was digested with *Cla*I to generate a full length cDNA with 18 bp 5' untranslated and 203 bp 3' untranslated, and this fragment was blunt ended using Klenow polymerase (Pharmacia Biotech) and a mixture of dNTPS and cloned into the *Sma*I site of pVL1392 to give the plasmid pVLC4A1.

pVLE26: The baculovirus transfer vector was constructed using the cDNA clone E-26 in vector pBluescript (SK-) (Pihlajaniemi *et al.*, J. Biol. Chem. 265:16922-16928 (1990)) and the polyhedrin-based transfer vector pVL1392

(Luckow *et al.*, Virology 170:31-39 (1989)). The cDNA clone E-26 encodes the  $\alpha 1$  chain of human type XIII collagen that is ligated into the *Eco*R1 site of pBS(SK-) (construct termed clone E-26). The E-26 clone is described in, for example, Pihlajaniemi *et al.*, J. Biol. Chem. 265:16922-16928 (1990). The cDNA E-26 was  
5 obtained from a  $\lambda$ gt11 cDNA library derived from human umbilical vein endothelial cells (Clontech), and the insert was released by digestion with *Eco*RI. This *Eco*RI fragment was ligated into the *Eco*RI site of pBR322 to give the clone E-26. Clone E-26 is digested with *Eco*R1 to generate the E-26 cDNA covering type XIII coding  
10 sequences. 123 bp 5' untranslated region and 117 bp 3' untranslated region are included, and this fragment is cloned into the *Eco*R1 site of pVL1392 to give the plasmid pVLE26.

pVLhuXIII: The baculovirus transfer vector was constructed using clone E-26  
15 (Pihlajaniemi *et al.*, J. Biol. Chem. 265:16922-16928 (1990)), genomic human type XIII collagen sequences (Tikka *et al.*, J. Biol. Chem. 266:17713-17719 (1991)) and the polyhedrin-based baculovirus transfer vector pVL1932 (Luckow *et al.*, Virology 170:31-39 (1989)). A clone called pBShuXIII was constructed and it contains the  
20 clone E-26 of the  $\alpha 1$  chain of human type XIII collagen with the 5' end of genomic human type XIII collagen covering nucleotides 1-272 from the type XIII collagen gene generated by PCR, in the *Not*I-*Eco*R1 site of pBS(SK-) to give the full-length cDNA of type XIII collagen (Tikka *et al.*, J. Biol. Chem. 266:17713-17719 (1991)).  
25 The 5' end of the genomic human type XIII collagen was generated using CL412 (a lambda clone isolated from a human genomic library (Clontech)) as the template and the PCR primers: 5' primer (5'-ATGCGGCCGCACGCGAGAGGATGGTAGC-3'), and 3' primer (5'-TAGCTGTCTCCATTGCTGCTC-3'). The 5'-PCR-primer included a new *Not*I restriction site preceding the type XIII sequences, which was  
30 used as well as a *Pst*I site between nucleotides 216 and 217 (Tikka *et al.*, J. Biol. Chem. 266:17713-17719 (1991)), for cloning the 5'-PCR-product into the clone E-26 digested with *Not*I and *Pst*I (Pihlajaniemi *et al.*, J. Biol. Chem. 265:16922-16928 (1990)). pBShuXIII is digested with *Not*I-*Eco*R1 to generate the full-length cDNA  
35 with 10 bp 5' untranslated region and 117 bp 3' untranslated region, and this

fragment is cloned into the *NotI*-*EcoRI* sites of pVL1392 to give the plasmid pVLhuXIII.

pVLmoXIII: The baculovirus transfer vector was constructed using the vector pBSmoXIII and the polyhedrin-based baculovirus transfer vector pVL1392, which is described in Luckow *et al.*, *Virology* 170:31-39 (1989). pBSmoXIII consists of the cDNA clone 689 in pBluescript encoding the  $\alpha 1$  chain of mouse type XIII collagen wherein the 5' end was generated by PCR and the 3' end by ligation of a fragment from the plasmid moC-2. Clone 689 is a cDNA derived from mouse spleen RNA as follows: total mouse spleen RNA is reacted with reverse transcriptase and the primer 5'-ACACACACAGGCCAGT-3'. The reverse transcriptase products are then used as a template for a first PCR reaction with the primers: 5' primer (5'-ATGAATTCGCCAGTCCCAGGTTAGAGGCA-3'), and 3' primer (5'-ATGAATTCAAGTTCTACTCGCGTAGGCGC-3'), and these products were used as a template for second PCR reaction with the primers: 5' primer (5'-ATGAATTCGTTCCAGCAGCCTTGGACTG GTAAGC-3'), and 3' primer (5'-ATGAATTCCCGAAGATGTCTCCAGGATGT- 3'). The PCR fragment covers nucleotides 466-969 from the cDNA sequences for mouse  $\alpha 1$  chain of type XIII collagen. cDNA clone GUT 219.2.4 was used as a template with the PCR primers: 5' primer (5'-ATAAGCTTGAATTCGAGGGCATGGTGGCGG-3'), and 3' primer (5'-CGAGGCCCGACGATGGACAT-3'). GUT 219.2.4 was obtained from a cDNA library derived from newborn mouse gut RNA using random hexamers as primers and the You-Prime-cDNA synthesis kit (Pharmacia) by probing with reverse transcriptase - PCR clones of the NC1 to NC4 domain of mouse type XIII collagen. These RT-PCR clones were obtained as follows: newborn mouse gut DNA was used as a template with the primers: 5' primer (5'-ACCTTTGGCCCTGGGGGCGCAGGGAGC-3'), and 3' primer (5'-AGGGAGAGAAAGGCGATGCTGGCA-3') to produce the M91 fragment. The M91 fragment was used to design primers for subsequent RT-PCR reactions in both the 5' and 3' directions using combinations of primers of mouse and human origin. Clone 689 was digested with *EcoRI* and *BamHI*, and this 557 bp *EcoRI*-*BamHI* fragment was ligated into the *EcoRI* and *BamHI* sites of pBluescript to give the

plasmid P40-5. Plasmid P40-5 was digested with *Hind*III and *Bbv*II and ligated with the 5'-PCR fragment digested with *Hind*III and *Bbv*II to give plasmid P40-1.

Plasmid P40-1 contains a clone with the translation initiation region and coding sequences up to the *Bam*HI site at nucleotide 1419.

5 The 3' end of the mouse type XIII collagen was added to plasmid P40-1 as follows. A mouse type XIII collagen cDNA, MOABCD.5 was used to obtain the 590 bp *Stu*I-*Sac*I fragment. MOABCD.5 was built from the cDNA clones GUT 229A, GUT 219.1.4, RG.6, and I8 to cover the coding sequences of mouse type XIII  
10 collagen except for the alternatively spliced exons 4A, 4B, 12, 13, and 33. Clones GUT 229A and GUT 219.1.4 are obtained from a cDNA library produced from newborn mouse gut RNA using random hexamers as primers and the You-Prime-cDNA synthesis kit (Pharmacia) by probing with reverse transcriptase -  
15 PCR clones of the NC1 to NC4 domain of mouse type XIII collagen. These RT-PCR clones were obtained as follows: newborn mouse gut DNA was used as a template with the primers: 5' primer (5'-ACCTTTGGCCCTGGGGGCGCAGGGAGC-3'), and 3' primer (5'-AGGGAGAGAAAGGCGATGCTGGCA-3') to produce the M91 fragment. The  
20 M91 fragment was used to design primers for subsequent RT-PCR reactions in both the 5' and 3' directions using combinations of primers of mouse and human origin. Clone RG.6 was obtained using 3'-RACE-PCR. The reverse transcriptase reaction was carried out using mouse gut poly(A+) RNA as the template and the primer  
25 (5'-GACTGC AGTCGACATCGATTTTTTTTTTTTTTTT-3'), followed by a first round of PCR using primers: MR-15 (5'-GCCTCCAGGAATGAAGGGAGAAGT-3'), and primer Tail (5'-GACTCGAGTCGACATCG-3'), and a second round of PCR using the primers:  
30 MR-1 (5'-GGGGGAGAGGGGGAAGAA-3'), and primer Tail (5'-GACTCGAGTCGACATCG-3'). These products are ligated to the vector pCRII (Invitrogen), and clone RG.6 was isolated from this library using the PCR clones of the NC1 to NC4 domain of mouse type XIII collagen as probes. Clone I8 was  
35 obtained using RT-PCR. The reverse transcription reaction used newborn mouse GUT RNA as template and oligo(dT)17 as a primer, followed by a first round of

PCR using primers: M-11 (5'-ATGCCATCGGAGGAGGCAG-3') and MR-13 (5'-GTTCCAGCAGCCTTGGACTGGTAAGC-3'), and a second round of PCR with the primers: MR-15 (5'-GCCTCCAGGAATGAAGGGAGAAGT-3') and MR-13 (5'-GTTCCAGCAGCCTTGGACTGGTAAGC-3'). These products were ligated into pCR1000 (Invitrogen) and clone I8 was obtained by probing with the M91 fragment. The GUT 229A, GUT 219.2.4, RG.6 and I8 inserts were liberated from their respective library vectors by using *NotI*. These *NotI* fragments were ligated into the *NotI* site of plasmid pBluescript to give the clones GUT 229A, GUT 219.1.4 and RG.6.

MOABCD.5 was constructed as follows: Clone pBluescript-GUT 219.1.4 was digested with *BamHI* and *EcoRI*, and the resulting 960 bp fragment was ligated into *BamHI/EcoRI* pBluescript-GUT 229A, to give clone MOAB.3. Plasmid pBluescript-I8 was digested with *StuI* and *HindIII* and the resulting 310 bp fragment was ligated into *StuI/HindIII* digested MOAB.3, to give the clone MOABC.5. The plasmid pBluescript-RG.6 was digested with *XbaI* and *HindIII*, and the resulting 250 bp fragment was ligated into *XbaI/HindIII* digested MOABC.5, to give the clone MOABCD.5

MOABCD.5 was digested with *StuI* and *SacI*, and the ensuing 673 bp *StuI-SacI* fragment was ligated into the *StuI* and *SacI* sites of clone 689 (plasmid P40) to give plasmid moC-2. Plasmid moC-2 was digested with *BamHI* and *SacI*, and this 1504 bp *BamHI-SacI* fragment is ligated to the *BamHI* and *SacI* sites of plasmid P40-1 to give the plasmid pBSmoXIII. pBSmoXIII is digested with *EcoRI* to generate a full-length type XIII collagen variant with seven base pairs 5' untranslated and 288 base pairs 3' untranslated, and this fragment was cloned into the *EcoRI* site of pVL1392 to give the plasmid pVLmoXIII. Another alternatively spliced full-length cDNA variant for the  $\alpha 1$  chain of mouse type XIII collagen was constructed and is termed pVLmoXIII(-E12/-E13). This construction is identical to pVLmoXIII, except that it lacks the sequence that encodes exon 12.

pVLC15A1: The baculovirus transfer vector was constructed using a PCR fragment covering nucleotides 14 to 1374 of type XV procollagen cDNA (Kivirikko *et al.*, *J. Biol. Chem.* 269: 4773-4779, (1994)). cDNAs for type XV procollagen

were made from human umbilical cord RNA using standard techniques described in, for example, Maniatis *et al.*, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, N.Y. (1989) and Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, N.Y. (1989). Using  
5 the cDNAs as a template, the PCR fragment covering nucleotides 14 to 1374 of type XV procollagen was made with the PCR primers: 5' primer (5'-GATATCACCTT CGTCCTCCGCTAAGCTC-3'), and 3' primer (5'-GAATTCTGGCC  
TCCACTTCCCCAGGCAT-3'). The PCR fragment contains an *EcoRV* linker  
10 sequence at the 5' end and an *EcoRI* linker sequence at the 3' end. The PCR fragment is digested with *EcoRV* and *EcoRI*, and ligated into the *EcoRV-EcoRI* sites of pBluescript (SK-). This construct was digested by *SphI* (cleaving in the PCR  
fragment at sequences corresponding to nucleotide 1355 of sequences presented in  
15 Kivirikko *et al.*, J. Biol. Chem. 269:4773-4779 (1994)) and *EcoRI* (digesting at the polylinker of pBluescript). An *SphI-EcoRI* fragment of clone SK5-3 covering  
nucleotides 1355-4330 in Kivirikko *et al.*, J. Biol. Chem. 269:4773-4779 (1994),  
was ligated with the above *SphI* and *EcoRI* digested construct resulting in construct  
20 pBShuXV. The clone SK5-3 was isolated from a  $\lambda$ gt11 cDNA library derived from human placenta (Clontech), the SK5-3 insert was released by digestion with *EcoRI*,  
and this insert was ligated into the *EcoRI* site of pBluescript (SK-) to give SK5-3. pBShuXV is digested with *EcoRV* (cleaving at pBluescript polylinker) and *HincII*  
(cleaving at nucleotide 4309 of type XV collagen cDNA sequences) to generate the  
25 full length cDNA for COL XV including 76 bp 5' untranslated region, and 53 bp 3' untranslated region, and this fragment is cloned in the *SmaI* site of pV11392  
(Luckow *et al.*, Virology 170:31-39 (1989)) to give the plasmid pVLCL5A1.

M18K: The baculovirus transfer vector was constructed using the  
30 polyhedron-based baculovirus transfer vector pVL 1393 (Invitrogen) and pBluescript SK M18kok.11 (pBsM18kok.11), which is made from the clones SXT-5B5, MM-103  
(Rehn *et al.*, J. Biol. Chem. 269:13929-13953 (1994)), and MM 21.3.

The cDNA SXT-5B5 was identified and cloned from a  $\lambda$ gt10 cDNA library  
35 made from mouse embryo (Clontech) as follows. The library was screened using a probe G2 for murine type XIII collagen to identify the clone ME-1. G2 had been

generated by RT-PCR using newborn mouse gut RNA as template, and primers: 5' primer MR-6 (5'-CCGGTGAGCCTGCTTGTCCT-3'), and 3' primer MR-11 (5'-ATGCCATCGGAGGAGGCAG-3'). The PCR product was ligated into the vector PCR-1000 (Invitrogen) and the construct was further digested with  
5 EcoRI-HindIII to give the probe G2. ME-1 covers 2.3-kB of the mouse  $\alpha 1$ (XVIII) mRNA (described in Rehn et al., Proc. Nat'l. Acad. Sci. USA 91: 4234- 4238 (1994)) and it was used to rescreen the mouse embryo library and identify the clone SXT-5B5, which was isolated by digesting with EcoRI, and ligating the SXT-5B5  
10 insert into the EcoRI site of pBluescript SK to give the plasmid pBs(SK)SXT-5B5 containing 540 bp extreme 5' sequence of the mouse  $\alpha 1$  (XVIII) chain clone SXT-5 (SXT-5 is described in Rehn et al., Proc. Nat'l. Acad. Sci. USA 91:4234-4238 (1994)).

15 The cDNA clone MM-103 was obtained from a  $\lambda$ gt10 cDNA library of poly(A) RNA from adult mouse liver (BALP/c strain) isolated by the guanidium thiocyanate method (Chomczynski et al., Anal. Biochem. 162:156-159 (1987)), followed by two rounds of oligo(dT)-cellulose chromatography. A cDNA library  
20 was constructed from this RNA using an oligo(dT) primer and the Time-Saver-cDNA synthesis kit (Pharmacia). This library is screened with a probe from ME-1, and the clone MM-103 was isolated by digesting with NotI. The insert MM-103 was ligated into the NotI site of pBluescript SK to give pBs(SK)MM-103. (Rehn et al., J. Biol. Chem. 269:13953(1994)).

25 The cDNA clone MM-21.3 was identified and cloned from a  $\lambda$ gt10 cDNA library, which had been made from adult mouse liver (BALB /c strain) poly(A) RNA using oligo(dT) method (described above, see also Rehn et al., J. Biol. Chem. 269:13929-13953 (1994)), by screening the library with probes SXT-5B5 and ME-1.

30 The clone was digested with NotI and ligated into NotI site of pBluescript SK to give the plasmid pBs(SK)MM-21.3, which covers nucleotides 360-2572 of the mouse  $\alpha 1$ (XVIII) chain (Rehn et al., Proc. Nat'l. Acad. Sci. 91:4234-4238(1994)).

The plasmid pBs(SK)SXT-5B5 was digested with EcoRI and the resulting 540  
35 bp fragment was further cloned into EcoRI-digested 5 kB fragment (insert + Bluescript) of plasmid pBs(SK)MM-21.3 to generate plasmid

pBsM18kc.AB.pBsM18kc.AB was digested with EcoRV and NsiI, resulting in a 2.5 kB fragment, and plasmid pBsMM 103 was digested with NsiI and NotI resulting in a 1.5 kB fragment. These two fragments were ligated into EcoRV-NotI- digested vector pBluescript to give plasmid pBsM18kok.11, which contains the full-length  
5 cDNA of the shortest variant of the  $\alpha 1$  chain of mouse type XVIII collagen (1315 amino acid residues) including 22 bp 5' untranslated region and 180 bp 3' untranslated region. pBsM18kok.11 was digested with EcoRV-NotI, and this fragment is cloned into the SmaI-NotI sites of pV1.1393 to give the plasmid M18K.

10 M18VA2K: The baculovirus transfer vector was constructed using the polyhedron-based baculovirus transfer vector pVL 1393 (Invitrogen), and pBsv2.5 which was built from cDNA clones PE17.24 (Rehn et al., J. Biol. Chem. 269:13929-13953 (1994)) and PX4.3 (Rehn et al., J. Biol. Chem. 269:13929-13953  
15 (1994)), and plasmid pBsM18kok.11 (described previously, see the construct M18K) to generate pBsM18VA2K.

The cDNA clone PE17.24 is isolated from a cDNA pool made from 18.5 day-old mouse embryo poly(a) RNA with the primer  
20 (5'-GATGGCAAATAGCACCC-3'). The cDNA from this synthesis are ligated into  $\lambda$ gt10 vectors, and the products are screened using a probe from ME-1. The clone PE17.24 was identified in this way and the insert was isolated by digesting with NotI, and the PE17.24 insert was ligated into the NotI site of pBluescript SK to give the pBs(SK)PE17.24.

25 Plasmid v2.5 was built from clones PE17.24 and PX4.3 by digesting the pBsPE17.24 with EcoRI and ligating into the resulting 4.8 kB fragment (insert + Bluescript) the EcoRI digested 90 bp fragment of PX4.3 to cover the long 764 residues form of the type XVIII collagen NCI domain.

30 To construct the full-length cDNA encoding the longest variant  $\alpha 1$ (XVIII) chain (1774 amino acid residues), including 3 bp 5' untranslated region and 180 bp 3' untranslated region, the plasmid v 2.5 was digested with Clal and the resulting 1.5 kB fragment was ligated into a Clal-digested 7.3 kB fragment (insert + Bluescript,  
35 the other Clal site in Bluescript) of pBsM18kok.11 (plasmid described previously, see construct M18K) resulting in the clone pBsM18VA2K. pBsM18VA2K was



further digested with EcoRV and NotI, and the resulting fragment was cloned into the *SmaI*-*NotI* sites of pVL1393 to give M18VA2K.

M18NC1: The baculovirus transfer vector was constructed using the cDNA clones SXT-5B5 (described above, see construct M18K) and ME-1 (Rehn et al., Proc. Nat'l. Acad. Sci. 91:4234-4238(1994)), and the polyhedron- based baculovirus transfer vector pVL 1393 (Invitrogen). SXT-5B5 was identified and cloned as described previously (see construct M18K). ME-1 covers 2.3 kB of the mouse  $\alpha$  (XVIII) mRNA (described in Rehn et al., Proc. Nat'l. Acad. Sci. 91:4234-4238 (1994)), and it encodes the N-terminal noncollagenous domain (NCI) of the shortest variant of the mouse type XVIII collagen  $\alpha$ 1 chain (characterizing and isolating of the clone ME-1 described previously, see construct M18K).

A stop codon is generated to the 3' end of the NCI domain by PCR, using ME-1 as template and the primers: 5' primer - T7, 17-mer primer (5'-AATACGACTCACTATAG-3'), and the 3' primer M18BacI (5'-GAAGGGGCTTGATAAATGAGGATCCAT-3') including an in-frame stop codon and a BamHI digestion site. The 400 bp PCR product was digested with EcoRI and BamHI and ligated to EcoRI-BamHI-digested pBluescript SK to give plasmid pBsNCII, pBsSXT-5B5 was digested with EcoRI and the resulting 540 bp fragment was further cloned into the EcoRI- digested pBsNCIL to give the plasmid pBsM18NCI, encoding the NCI domain and 22 bp of 5' untranslated sequences. pBsM18NCI is digested with EcoRV-NotI and the resulting fragment is cloned into the *SmaI*-*NotI* sites of the pVI 1393 to give the plasmid M18NCI.

M18VA2N: The baculovirus transfer vector was constructed using the polyhedron-based baculovirus transfer vector pVL1393 (Invitrogen), and the plasmid pBsM18NCI (described previously, see construct M18NCI), and the plasmid pBsV2.5 (see the construct M18VA2K). Plasmid pBsV2.5 was digested with ClaI and the resulting 1.5 kB fragment was cloned into the ClaI-digested 4.8 kB fragment of pBsM18NCI to generate the plasmid pBsM18VA2.3 encoding the longest variant aminoterminal noncollagenous domain (NCI-764) of type XVIII collagen  $\alpha$ 1 chain. pBsM18VA2.3 is digested with EcoRV-NotI and the resulting fragment is cloned into the *SmaI*-*NotI* sites of pVL1393 to give the plasmid M18VA2N.

M18C: The baculovirus transfer vector was constructed using the vector pBluescript (SK)MM-103 (described previously, see the construct M18K) and the polyhedron-based baculovirus transfer vector pVL 1393 (Invitrogen).

5 pBluescript(SK)MM-103 encodes the cDNA for the C-terminus of the  $\alpha 1$  chain of mouse type XVIII collagen in the NotI site of pBluescript SK. pBs(SK)MM-103 was digested with EcoRI-NotI which generates a cDNA fragment covering nucleotides 2802-4080 (see, Rehn et al., J. Biol. Chem. 269:13929-13953 (1994)) with a translation initiation codon at nucleotides 3010-3012 corresponding to the C-terminal  
10 noncollagenous domain (amino acid residues 997-1315) with 180 bp of the 3' untranslated region. This fragment is cloned into the EcoRI-NotI sites of the pVL1393 to give the plasmid M18C.

15                    2.     *Construction of Recombinant Vectors Containing Collagen Modifying Enzymes.*

pVL $\beta$ : The baculovirus transfer vector was constructed using the polyhedrin-based baculovirus transfer vector pVL 1392, and the vector pBS(SK-)S138 which contains the full length cDNA for the  $\beta$ -subunit of human  
20 prolyl 4- hydroxylase in the *EcoRI* site (Pihlajaniemi *et al.*, EMBO, J. 6:643 (1987)). The  $\beta$ -subunit clone HB-95 was obtained from a human hepatoma  $\lambda$ gt11 cDNA expression library screened with purified antibodies against human prolyl 4-hydroxylase. The vector pBS(SK-)S138 was constructed by identifying the clone  
25 S138 (human prolyl 4- hydroxylase  $\beta$ -subunit) from a  $\lambda$ gt11 library derived from human placenta (Clontech) using HB-95 as a probe for the  $\beta$ -subunit of human prolyl 4-hydroxylase, releasing the insert from the identified  $\lambda$ gt11 clone with *EcoRI*, and inserting the *EcoRI* fragment into the *EcoRI* site of pBS(SK-) (Stratagene) to give  
30 pBS(SK-)S138.

pSB(SK-)S138 was digested with *EcoRI*-*Bam*HI to generate the full length cDNA plus 44 bp 5' untranslated and 207 bp 3' untranslated, and this fragment was cloned into the *EcoRI*-*Bam*HI sites of pVL1392 (Vuori *et al.*, Proc. Natl. Acad. Sci. USA 89:7467-7470 (1992)) to give the plasmid pVL $\beta$   
35

pVL $\alpha$ : The baculovirus transfer vector was constructed using the vector pBS(SK-)PA59 which contains the full length cDNA for human prolyl 4-hydroxylase

$\alpha$ -subunit in the *Sma*I site (Helakoski *et al.*, Proc. Nat'l. Acad. Sci. USA 86:4392-4396 (1989)) and the polyhedrin-based baculovirus transfer vector pVL 1392. The cone PA59 (human prolyl 4-hydroxylase  $\beta$ -subunit) was obtained as follows. An oligonucleotide mixture which encodes a peptide  
5 (Gln-Val-Ala-Asn-Tyr-Gly) from the  $\alpha$ -subunit of prolyl 4-hydroxylase was used to screen a cDNA library from HT 1080 cells. One positive clone, HTA-2, was obtained, and a 36-mer oligonucleotide derived from the HTA-2 sequence (nucleotides 1430-1465 of the  $\alpha$ -subunit) was used to screen a human placenta  $\lambda$ gt11  
10 library (Clontech). Two positive clones, PA-11 and PA-15, were isolated, and the full length clone, PA59, was obtained by rescreening the placenta library with these clones.

The vector pBS(SK-)PA59 was constructed by releasing the  $\lambda$ gt11 insert from  
15 the clone PA59 by digestion with *Hin*P1 and *Acc*I, blunt ending the PA59 fragment with Klenow (Pharmacia Biotech), and cloning the blunt ended PA59 fragment into the *Sma*I site of pBS(SK-) (Stratagene) to give pBS(SK-)PA59. pBS(SK-)PA59 was digested with *Pst*I and *Bam*HI to generate *Pst*I-*Pst*I and *Pst*I-*Bam*HI fragments  
20 containing the full length cDNA plus 61 bp 5' untranslated region, and 551 bp 3' untranslated region, and these fragments are cloned into the *Pst*I- *Bam*HI sites of pVL1392 (Vuori *et al.*, Proc. Natl. Acad. Sci. USA 89:7467-7470 (1992)) to give the plasmid pVL $\alpha$ .

p2Bac $\beta$ : pBS(KS-)S138 was constructed by digesting pBS(SK-)S138 with  
25 *Eco*RI to release the S138 clone, and then inserting the S138 fragment into the *Eco*RI site of pBS(KS-)S138. pBS(KS-)S138 was digested with *Bam*HI to give the full length  $\beta$ -subunit of human prolyl 4- hydroxylase including 44 bp 5' untranslated region and 207 bp 3' untranslated region. This fragment was cloned into the *Bam*HI  
30 site of p2Bac to give p2Bac $\beta$ .

pBS(SK-)PA59 was mutated by PCR to place a *Not*I site 46 bp upstream of the initiation codon for the  $\alpha$ -subunit of prolyl 4-hydroxylase to give the plasmid pBS(SK-)PA59/5'UT*Not*I as follows. The plasmid pBS(SK-)PA59 and the primers:  
35 5' primer (5'-GCCCTCGCGGCCGCTTTCCAGGT-3'), and 3' primer (5'-TGACATATCCTTAAGGACCAGTTC-3'), are used in a first PCR reaction,

followed by a second PCR reaction using pBS(SK-)PA59 and the primers: 5' primer (5'- CGAGGTATCGATAAGCTTG-3'), and 3' primer (fragment from the first PCR reaction). The second PCR product is digested with *Clal* and *Afl*III, and ligated into pBS(SK-)PA59 to generate pBS(SK-)PA59/5'UTNotI. pBS(SK-)PA59/5'UTNotI is digested with *NotI* to generate a fragment with the full length  $\alpha$ -subunit of prolyl 4-hydroxylase including 46 bp 5' untranslated region and 551 bp 3' untranslated region. This fragment is cloned into the *NotI* site of p2Bac $\beta$  to give the plasmid p2Bac $\beta$ .

3. *Expression of Recombinant Collagen Genes in Insect Cells with Prolyl-4-Hydroxylase.*

Recombinant human collagens I, II, III, IV, XIII, XV, and XVIII have been expressed in insect cells by means of baculovirus expression vectors.

Expression of Collagen Type III. pVLC3A1 is a recombinant expression vector encoding the full pro $\alpha$ 1 chain of human type III collagen. Similar baculovirus expression vectors pVL $\alpha$ , pVL $\beta$ , and p2Bac $\beta$  were created for the expression of human prolyl 4-hydroxylase in insect cells. The constructs were transfected in various combinations into insect cells using a BaculoGold<sup>®</sup> transfection kit (Pharmingen).

Insect cells (Sf9 or High Five, Invitrogen) were cultured in TNM-FH medium (Sigma) supplemented with 10% fetal bovine serum (BioClear) or in a serum-free HyQ CCM3 medium (HyClone) either as monolayers or in suspension in shaker flasks at 27°C. To produce recombinant proteins, insect cells seeded at a density 5-6 x 10<sup>5</sup>/ml were infected at a multiplicity of 5-10 with the recombinant virus and at a multiplicity of 1 with the viruses for the ( subunit and ( subunit of human prolyl 4-hydroxylase (Vuori *et al.*, Proc. Natl. Acad. Sci. USA 89:7467-7470 (1992)). Ascorbate (80  $\mu$ g/ml) was added daily to the culture medium. The cells were harvested 48- 120 h after infection, washed with a solution of 0.15 M NaCl and 0.02 M phosphate, pH 7.4, homogenized in a 0.3 M NaCl, 0.2% Triton X-100 and 0.07 M Tris buffer, pH 7.4, and centrifuged at 10,000 x g for 20 min. The remaining cell pellet that was insoluble in the homogenization buffer was further solubilized in

1% SDS and analyzed by SDS-PAGE. The cell culture medium was concentrated 10 times in an ultrafiltration cell (Cmicon) with a PM-100 membrane. Aliquots of the supernatants of the cell homogenates and the concentrated cell culture medium were analyzed by denaturing SDS-PAGE, followed by staining with Coomassie Brilliant Blue or Western blotting with an antibody to the N-propeptide of human type III procollagen.

More specifically, Sf9 and High Five cells were infected with a recombinant baculovirus coding for the pro $\alpha$ 1 (III) chains, harvested 72 h after infection, homogenized in a buffer containing 0.2% Triton X-100 and centrifuged. Aliquots of the Triton X-100 soluble protein fraction and the concentrated cell culture medium were then analyzed either without pepsin treatment or after treatment with pepsin for 1h at 22°C. The samples were electrophoresed on 8% SDS-PAGE and analyzed by Coomassie staining in A and by Western blotting using an antibody to the N-propeptide of human type III procollagen in B. As set forth in Figure 6, Lane 1 sets forth molecular weight markers; lanes 2-3, cell extracts; and lanes 4-5, media from Sf9 cell cultures; lanes 6-7, cell extracts; and lanes 8-9, media from High Five cell cultures. Samples in the odd numbered lanes were digested with pepsin. Because the antibody used in the Western blotting reacts only with the N-propeptide of type III procollagen, it does not recognize pepsin digested samples. The arrows indicate the pro $\alpha$ 1 (III) and  $\alpha$ 1 (III) chains.

Other aliquots were studied by a radioimmuno assay for the trimeric N-propeptide of human type III procollagen (Farmos Diagnostica) and a colorimetric method for 4-hydroxyproline (Kivirikko *et al.*, Anal. Biochem. 19:249-255 (1967)). Still further aliquots were digested with pepsin for 1h at 22°C (Bruckner *et al.*, Anal. Biochem. 110:360-368 (1981)), and the thermal stability of the pepsin-resistant recombinant type III collagen was measured by rapid digestion with a mixture of trypsin and chymotrypsin.

The expression level of pro $\alpha$ 1 (III) could be seen by Western blotting in samples of the Triton X-100 soluble proteins (Fig. 6B, lanes 2 and 6) and cell culture media (Fig. 6B, lanes 4 and 8) in both Sf9 and High Five cells. After the pepsin digestion the (I chains of type III collagen were seen in the High Five cells in

the Coomassie stained gel (Fig. 6A, lane 7). The pepsin resistant (I(III) chains were not detected in the Western blot (Fig. 6B, lanes 3, 5, 7 and 9) since the antibody used reacts only with the N-propeptides of the pro $\alpha$ 1(III) chains, which were

5 apparently digested by pepsin.

Sf9 and High Five cells were infected with the virus coding for the pro $\alpha$ 1 (III) chains either with or without viruses coding for the two types of subunit of prolyl 4-hydroxylase (TABLE III). The expression level of total type III procollagen was measured with a radioimmuno assay for the trimeric N-propeptide, and the amount of

10 4-hydroxyproline formed in the cells was determined by a colorimetric assay. Both values were used to calculate the amount of type III collagen produced by assuming that all the pro $\alpha$ 1 (III) chains formed triple-helical molecules and that all the hydroxylatable proline residues in the pro $\alpha$ 1 (III) chains had been converted to 4-

15 hydroxyproline. Based on the known structure of type III procollagen and the amount of 4-hydroxyproline in type III collagen, the amount of type III collagen in the samples was calculated by multiplying the N-propeptide values obtain by 7 and the 4-hydroxyproline values by 8. All measurements were made 72 h after the

20 infection.

A considerable variation was found in the values obtained in different experiments as shown in TABLE II. Notwithstanding this variation, TABLE II provides: First, the amount of 4-hydroxyproline formed was in all experiments distinctly higher in cells infected with the prolyl 4-hydroxylase-coding viruses than in

25 their absence. Second, the expression level obtained in High Five cells was consistently higher than that obtained in Sf9 cells. Third, in cells coinfectd with the prolyl 4-hydroxylase-coding viruses the level of type III collagen produced was always higher when calculated from the 4-hydroxyproline values than from the

30 radioimmuno assay values, suggesting either that some of the N-propeptides of type III procollagen were degraded or that some of the fully 4-hydroxylated pro $\alpha$ 1 (III) chains remained nontriple-helical. The highest type III collagen expression values were in the High Five cells that also expressed prolyl 4-hydroxylase, the amount of

35 cellular type III collagen in these cells being about 41-81  $\mu$ g/5 x 10<sup>6</sup> cells (TABLE III). The amount of type III collagen secreted into the culture medium, when

measured with the radioimmuno assay, was about 25-50% of total in Sf9 cells and about 10-30% of total in High Five cells.

Experiments were also performed in which High Five cells were grown in suspension in shaker flasks. A similar effect of prolyl 4-hydroxylase-coding viruses was seen in these experiments as above. The highest expression levels found in such experiments have ranged up to about 40 mg of type III collagen produced per liter of culture in 72 h, about 80-90% of the collagen produced being found in the cell pellet, and 10-20% in the medium.

10

TABLE III

**PROLYL 4-HYDROXYLASE ACTIVITY OF TRITON X-100  
EXTRACTS FROM INSECT CELLS EXPRESSING PRO $\alpha$ 1 CHAINS  
OF HUMAN TYPE III PROCOLLAGEN WITH OR WITHOUT THE  
A AND B SUBUNITS OF PROLYL 4-HYDROXYLASE.**

15

20	Cells and recombinant polypeptides expressed	Prolyl 4-hydroxylase activity
		dpm/10 $\mu$ l
	<i>High Five cells</i>	
	None	Ê480
	Pro $\alpha$ 1 (III) chains	Ê500
25	Pro $\alpha$ 1 (III) chains and $\alpha$ and $\beta$ subunits	4810
	Sf9 cells	
	None	Ê150
	Pro $\alpha$ 1 (III) chains	ÊÊ60
30	Pro $\alpha$ 1 (III) chains and $\alpha$ and $\beta$ subunits	3360

The cells expressed either no recombinant polypeptide or only the pro $\alpha$ 1 (III) chains or the latter plus the  $\alpha$  and  $\beta$  subunits of prolyl 4-hydroxylase. The analysis was performed 72 h after the infection.

35

The values are given as dpm/10  $\mu$ l of the Triton extract, mean of duplicate values obtained in three experiments for High Five cells, and mean of duplicate values in one experiment for Sf9 cells.

5

Expression of Collagen Types I and II. Baculovirus expression vectors pVLC1A1 and pVLC1A2 were created for the expression of the pro $\alpha$ 1 chain and the pro $\alpha$ 2 chain of human collagen I, and pVLC3A15'UT/C2A1 was created for the expression of the pro $\alpha$ 1 chain of human collagen II.

10

Unless otherwise specified, insect cells were cultured, and recombinant collagen produced following the procedures supra.

The expression level of pro $\alpha$ 1 (I), and pro $\alpha$ 1 (I) and pro $\alpha$ 2 (I) in the presence of prolyl 4-hydroxylase, and following pepsin digestion of the supernatants from cell homogenates could be seen in silver-stained 5% SDS-PAGE. See Figure 7, lanes (DIA 1). The silver-stained SDS PAGE revealed the formation of triple-helical procollagen I in these cells. Homotrimeric collagen can be separated from heterotrimeric collagen I on a metal chelate affinity column through the use of a histidine-tag to the C-terminal domain of the pro $\alpha$ 2 chain.

20

The expression level of pro $\alpha$ 1 (II) in the presence of prolyl 4-hydroxylase could be seen in coomassie stained 5% SDS PAGE. See Figure 8 (wherein lane 1 depicts the expression of a homotrimer of type I collagen; lane 2 is a standard sample of type II procollagen; lane 6 is a standard sample of type III procollagen; and lanes 3-5 compare three different constructs of human type II procollagen containing varying amounts of human procollagen type III. Lane 3 is type II procollagen with the C-terminal end of type III procollagen; lane 4 is type II procollagen with the N-terminal non-collagenous region from type III procollagen; and lane 5 is type II procollagen with the N- and C-terminal regions of type III procollagen).

25

30

Several baculovirus vectors for the expression of human type II collagen were constructed. In one of these vectors, the 5' untranslated region of human type II collagen was replaced with human type III collagen 5' untranslated region. In another vector, the entire human type II collagen gene was expressed. In another

35



insect expression vector, the N-propeptide of type II collagen was replaced with an N-propeptide of type III collagen. All three of those vectors were found to express human type II collagen in varying levels. Expression was detected by Coomassie Blue stain SDS-PAGE and by Western blot analysis.

Expression of Collagen Types IV, XIII, and XVIII. pVLC4A1 is a recombinant baculovirus expression vector encoding the pro $\alpha$ 1 chain of human collagen IV. pVLhuXIII is a recombinant baculovirus vector encoding the pro $\alpha$ 1 chain of human collagen XIII. pVLC15A1 is a recombinant expression vector encoding the pro $\alpha$ 1 chain of human collagen XV. M18K and M18VA2K are recombinant expression vectors encoding two variants of the pro $\alpha$ 1 chain of human collagen type XVIII.

Unless otherwise specified, insect cells were cultured and recombinant collagen produced following the procedures *supra*. pVLC4A1, pVLhuXIII, pVLC15A1, M18K, and M18VA2K have been transformed into insect cells, and the recombinant collagens have been successfully expressed.

#### 4. Purification And Analysis Of Recombinant Collagen.

Purification of Recombinant Type III Collagen. The properties of the purified human type III collagen produced in insect cells were found to be very similar to those of the type III collagen extracted from various tissues (Kielty *et al.*, Connective Tissue and Its Heritable Disorders: Molecular, Genetic and Medical Aspects pp. 103-147 (1993); Kivirikko, Ann. Med. 25:113-125 (1993); van der Rest *et al.*, Adv. Mol. Cell. Biol. 6:1-67 (1993); Brewton *et al.*, Extracellular Matrix Assembly and Structure pp. 129-170 (1994); Pihlajaniemi *et al.*, Prog. Nucleic Acid Res. Mol. Biol. 50:225-262 (1995); Prockop *et al.*, Annu. Rev. Biochem. 64:403-434 (1995)). In particular, the content of 4-hydroxyproline and the Tm of the triple helices, when determined by CD spectra, were found to be virtually identical to those of the authentic type III collagen. The content of hydroxylysine in the recombinant collagen was found to be about one-half of that of type III collagen

extracted from various tissues, indicating that insect cells must have a considerable level of lysyl hydroxylase activity.

Insect cells expressing the recombinant type III procollagen were washed with a solution of 0.15 M NaCl and 0.02 M phosphate, pH 7.4, homogenized in a cold  
5 0.2 M NaCl, 0.1% Triton X-100 and 0.05 M Tris buffer, pH 7.4 (20 x 10<sup>6</sup> cells/ml), incubated on ice for 30 min, and centrifuged at 16,000 x g for 30 min. Unless otherwise mentioned, all the following steps were performed at 4°C. The supernatant was chromatographed on a DEAE cellulose column (DE-52, Whatman)  
10 equilibrated and eluted with a 0.2 M NaCl and 0.05 M Tris buffer, pH 7.4, the void volume being collected. The pH of the sample was lowered to 2.0-2.5, and the sample was digested with a final concentration of 150 µg/ml of pepsin for 1 h at 22°C. Pepsin was irreversibly inactivated by neutralization of the sample followed  
15 by an overnight incubation on ice. The recombinant type III collagen was precipitated by adding solid NaCl to a final concentration of 2 M and centrifugation at 16,000 x g for 1 h. The pellet was dissolved in a 0.5 M NaCl, 0.5 M urea, and 0.05 M Tris buffer, pH 7.4, for 1 day, and the sample was digested with pepsin as  
20 above for a second time. The sample was then chromatographed on a Sephacryl HR-500 gel filtration column (Pharmacia), eluted with a solution of 0.2 M NaCl and 0.05 M Tris, pH 7.4, dialyzed against 0.1 M acetic acid and lyophilized.

Type III procollagen was expressed in High Five cells cultured either as monolayers or in suspension in shaker flasks. The cells were harvested 72 h after  
25 infection, homogenized in a buffer containing 0.1% Triton X-100 and centrifuged, and the supernatant of the cell homogenate was passed through a DEAE cellulose column to remove nucleic acids. The flow through fractions containing the type III procollagen were pooled and digested with pepsin. This converted the type III  
30 procollagen to type III collagen and digested most of the noncollagenous proteins. The type III collagen was then concentrated by salt precipitation, solubilized and treated with pepsin as above. The type III collagen was finally separated from pepsin and other remaining contaminants by gel filtration on a Sephacryl S 500-HR  
35 column. The fractions containing the type III collagen were pooled, dialyzed and lyophilized.

The purified type III collagen was analyzed by 5% SDS-PAGE under reducing (Figure 9, lane 2) and nonreducing (Figure 9, lane 3) conditions. No contaminants were seen in the Coomassie stained gel and the type III collagen (I chains were disulfide-bonded. Amino acid and CD spectrum analysis were performed on the purified type III collagen. The amino acid composition of the recombinant type III obtained corresponded well with the amino acid composition reported for human type III collagen. The only exception was the amount of hydroxylysine, which was 3 residues/1000 amino acids in the recombinant type III collagen instead of 5/1000 amino acids in the authentic human type III collagen. The melting temperature of the recombinant type III collagen determined by CD spectrum analysis was 40°C.

The High Five cells gave consistently higher production rates than Sf9 cells, the highest production rates seen in High Five cells cultured in monolayers ranging up to about 80 µg of cellular recombinant human type III collagen/5 x 10<sup>6</sup> cells, which corresponds to about 120 µg of type III procollagen. When the High Five cells were cultured in suspension in shaker flasks, the highest amount of cellular type III collagen produced ranged up to about 40 mg/l, corresponding to about 60 mg/l of type III procollagen.

Conformational Integrity of the Recombinant Type III Collagen. Association of the pro $\alpha$ 1 (III) chains into trimers was studied by using SDS-PAGE analysis under nonreducing conditions. High Five cells were coinfecting with viruses coding for the pro $\alpha$ 1 (III) chains and the ( and ( subunits of human prolyl 4-hydroxylase. The cells were harvested 72 h after infection, homogenized in a buffer containing 0.2% Triton X-100, centrifuged, and the remaining cell pellets were further solubilized in 1% SDS. Aliquots of the Triton soluble proteins were treated with pepsin for 1 h at 22°C. Essentially all the pro $\alpha$ 1 (III) chains synthesized were found as disulfide-bonded trimers based on the disappearance of a protein band of a high molecular weight (Figure 10, lane 2). After pepsin digestion the band corresponding to the recombinant type III procollagen was converted to a band corresponding to type III collagen, and the protein remained in the form of the trimer, thus indicating

the existence of disulfide bonds between the (I (III) chains (Figure 10, lane 3).

Virtually all the type III procollagen expressed was soluble in the Triton

X-100-containing homogenization buffer, as no band corresponding to type III

5 procollagen was seen in the Triton X-100-insoluble, SDS-soluble fraction (Figure 10, lane 4).

The thermal stability of the type III collagen expressed under different cell culture conditions was studied by using digestion with a mixture of trypsin and chymotrypsin after heating to various temperatures (Bruckner, *et al.*, Anal. Biochem. 10 110:360-368 (1981)). High Five cells were infected with viruses coding for the pro $\alpha$ 1 (III) chains and the ( and ( subunits of human prolyl 4-hydroxylase. The cells were harvested 72 h after infection, homogenized in a buffer containing 0.2% Triton X-100 and centrifuged. In these experiments, ascorbate was either added daily to the 15 cell culture medium as usual or omitted during the infection. The Triton X-100 soluble proteins were first digested with pepsin for 1 h at 22°C to convert type III procollagen to type III collagen (Pihlajaniemi *et al.*, EMBO J. 6:643-649 (1987)), and the trypsin/chymotrypsin digestion was then performed for aliquots of the 20 pepsin-treated samples. The samples were then electrophoresed on 8% SDS-PAGE and analyzed by Coomassie staining. Figures 11A-11D provide the results of this thermal stability for a variety of collagen products. As set forth in panel A, the cells were infected only with the virus coding for the pro $\alpha$ 1 (III) chains, and ascorbate was omitted from the culture medium; panel B, the cells were infected only with the virus 25 coding for the pro $\alpha$ 1 (III) chains, and ascorbate was present in the culture medium as usually; panel C, the cells were coinfectd with viruses coding for the pro $\alpha$ 1 (III) chains, and the ( and ( subunits of prolyl 4-hydroxylase, but ascorbate was omitted from the culture medium; and panel D, the cells were infected with the three viruses, 30 and ascorbate was present in the culture medium. Lane P shows a sample digested with pepsin without subsequent trypsin/chymotrypsin digestion, lanes 27-42 show samples treated with the trypsin/chymotrypsin mixture at the temperatures indicated. The arrows show the position of the (I (III) chains. As evidenced by these results, 35 when the pro $\alpha$ 1 (III) chains were expressed without the presence of prolyl 4-hydroxylase and ascorbate, the Tm of type III collagen was found to be at about

32-34°C (Figure 11A). The presence of either ascorbate or prolyl 4-hydroxylase without the other had virtually no increasing effect on the thermal stability (Figure 11B and 11C). In contrast, when the pro $\alpha$ 1 (III) chains were produced in the presence of both prolyl 4-hydroxylase and ascorbate, the T<sub>m</sub> of type III collagen was increased considerably, being at about 38-40°C (Figure 11D).

Purification and analysis of Collagen Types I and II. Collagens types I and II were purified as described *supra*. The recombinant type II human collagen expressed from the recombinant insect cells was found to exhibit resistance to trypsin and chymotrypsin digestion. These protease digestion experiments indicated that triple helical type II human collagen was formed in the recombinant insect cells.

The thermal stability of the recombinant type II human collagen expressed from the recombinant insect cells was measured and compared with native type I human collagen. These results indicated that the recombinant type II collagen had a triple helical structure. The T<sub>m</sub> of the recombinant type II collagen was up to about 40°C.

20

**A. Example 11: Expression of Recombinant Collagen Genes in Yeast Cells Expressing Recombinant Genes for Prolyl 4-Hydroxylase**

**I. Construction of Recombinant Vectors Containing Collagen Genes.**

pPIC9ColIII. This plasmid contains the human Col III gene joined to the  $\alpha$ -mating factor secretion signal ( $\alpha$ -MFSS) (and containing a deletion of the native human secretion signal).

The 3' end of the COL III gene was synthesized by PCR from the 4195 bp downstream (*Eco*RI site) of the translation initiation codon to the stop codon (4401 bp) using pBluescript SM38 as a template and the PCR primers: 5' primer (5'-GAAGGTGAATTCAAGGCTGA-3'), and 3' primer (5'-GCGTCTAGAGCGGCCGCTTATAAAAAGCAAACAGGGCC-3'). *Not*I and *Xba*I sites were created in the 3' end of the PCR fragment. The PCR fragment was digested with *Eco*RI and *Xba*I and cloned into the *Eco*RI and *Xba*I sites of pBluescript-SM38 (pBS-SM38 is derived from sequences presented in Ala-Kokko *et al.* Biochem. J. 260: 509-516

(1989)), and GenBank accession number X14420) to give the plasmid pBluescript-SM38/B.

The 5' end of the Col III gene was synthesized from 73 bp downstream of the translation initiation codon to 176 bp (*Bam*HI site) by PCR (for sequences, see  
5 Ala-Kokko *et al.*, Biochem., J. 260:509-516 (1989)) using pBluescript SM38 as the template and the PCR primers: 5' primer (5'-GCGATCGATGC  
GGCCGCGCAGGAAGCTGTTGAAGGAGG-3'), and 3' primer (5'-GAGAA  
CGGATCCTGAGTCAC-3'). *Cla*I and *Not*I sites were created in the 5' end of the  
10 PCR fragment. pBluescript-SM38/B was digested with *Cla*I and *Bam*HI, and the fragments from this digest and the 5' PCR fragment were ligated with T4 ligase to give the plasmid pBluescript-SM38/11.

pBluescript-SM38/11 was digested by *Not*I and the *Not*I-*Not*I collagen  
15 fragment (73-4401 bp) was cloned in frame with the  $\alpha$ -factor signal sequence in the yeast expression vector pPIC9 (Invitrogen) to give the plasmid pPIC9COLIII.

pHIL-D2/colIII. The 3' end of the COL III gene was synthesized by PCR from the 4195 bp downstream (*Eco*RI site) of the translation initiation codon to the  
20 stop codon (4401 bp) using pBluescript-SM38 as the template DNA and the primers: 5' primer (5'-GAAGGTGAATTCAAGGCTGA-3'), and the 3' primer (5'-  
GCGTCTAGATTATAAAAAGCAAACAGGGCC-3'). An *Xba*I site was created in the 3' end of the PCR fragment. pBluescript-C3A1 was digested with *Eco*RI and  
25 *Xba*I and the large fragment isolated, and the 3' PCR fragment is digested with *Eco*RI and *Xba*I. These two fragments and the digested pBluescript-C3A1 vector are ligated with T4 ligase to give pBluescript-C3A1/10. A *Bgl*II site was created 16 bp upstream of the translation initiation codon in pBluescript- C3A1/10 and the *Bgl*II -  
30 *Xba*I fragment from pBluescript-C3A1/10, containing collagen sequences from (nucleotides - 16 to 4401) is ligated into the *Eco*RI site of pHIL-D2 (Invitrogen) to give plasmid PHIL-D2/colIII.

pAO815 $\beta$ . pYM25 was digested with *Hpa*I and the fragment containing the ARG4 gene of *Saccharomyces cerevisiae* was isolated and cloned into the *Eco*RV  
35 sites of pAO815 (Invitrogen) replacing the HIS4 gene with ARG4, to give the plasmid pARG815.

A cDNA of the  $\beta$  subunit of human prolyl 4-hydroxylase (Vuori *et al.*, Proc. Nat'l. Acad. Sci. USA 89:7467-7470 (1992)) was synthesized by PCR from the translation initiation codon to the stop codon, and *EcoRI* sites were created in the 5' and 3' ends of the PCR fragment. pVL1392/\_HDEL (Vuori *et al.*, 1992, EMBO J. 11:4213- 4217) was used as the template DNA with the primers: 5' primer (5'-GGCGAATTCATGCTGCGCCGCGCTCTGCT-3'), and 3' primer (5'-GCGGAA TTCTTACAGTTCATCGTGACAGC-3'). This PCR fragment was digested with *EcoRI* and cloned into pBluescript SK, to give pBluescript SK $\beta$ /20. pBluescript SK $\beta$ /20 was digested with *EcoRI* and this fragment was cloned into the *EcoRI* site of pAO815 (Invitrogen), to give the plasmid pAO815 $\beta$  which has a single expression cassette for the  $\beta$ -subunit of prolyl 4-hydroxylase.

pARG815 $\alpha$ . The 5' end of the  $\alpha$ -subunit of prolyl 4-hydroxylase was synthesized by PCR from the translation initiation codon to 689 bp downstream (*HindIII* site), and *HindIII* and *SmaI* sites were created in the 5' end of the fragment. pBS(SK-)/PA59 was used as the template DNA with the primers: 5' primer (5'-GCGAAGCTTCCC GGGATGATCTGGTATATATTA-3'), and 3' primer (5'-GGATCTAGTTCAAGAAGCTT-3'). pA-59 (Vuori *et al.*, Proc. Nat'l. Acad. Sci. USA 89:7467-7470 (1992)) was digested with *HindIII* and the large fragment was isolated and ligated with the 5' PCR fragment to give pA-59/15.

The 3' end of the  $\alpha$ -subunit was synthesized by PCR from 1373 bp (*PstI* site) downstream of the translation initiation codon to the translation stop codon, and *SmaI* and *BamHI* sites were created in the 3' end of the fragment. pBS(SK-)/PA59 was used as the template DNA with the primers: 5' primer (5'-AGTGATGTGTCTGCAGGAGGAGC- 3'), and 3' primer (5'-GCGGGATCCCCCGGGTCATTCCAATTCTGACAACG-3'). pA-59/15 was digested with *PstI* and *BamHI*, and the large fragment was isolated, and ligated with the 3' PCR fragment to give pA-59/3. pA-59/3 was digested with *SmaI* and the *SmaI*-*SmaI*  $\alpha$ -subunit fragment was cloned into the *EcoRI* site of pARG815, to give pARG815 $\alpha$ .

35

pARG815 $\alpha\beta$ . pAO815 $\beta$  was digested with *Bgl*II and *Bam*HI to excise the expression cassette, and the expression cassette is cloned into the *Bam*HI site of pARG815 $\alpha$  to give the vector pARG815 $\alpha\beta$ .

5 pAO815 $\beta\beta$  - is similar to pAO815 $\alpha\beta$ , but contains two cassettes of the  $\beta$  subunit of the human prolyl 4-hydroxylase gene. pAO815 $\beta$  was digested with *Bgl*II and *Bam*HI to excise the expression cassette, and the expression cassette is cloned into the *Bam*HI site of pARG815 $\alpha\beta$  to give the vector pARG815 $\alpha\beta\beta$ .

10 The  $\beta$ -subunit without its signal sequence was synthesized by PCR from 52 bp downstream of the translation initiation codon to the translation stop codon. *Eco*RI restriction sites were created in 5' and 3' ends. This PCR fragment was cloned into the *Eco*RI site of pSP72 (Promega).

## 15 2. *Expression of Recombinant Collagen Genes in Yeast Cells with Prolyl-4-Hydroxylase.*

*Pichia pastoris* host strain GS200 his4 arg4 was stably transformed with combinations of the plasmid described *supra* and related plasmids to produce the following recombinant strains.

20 *P. pastoris* Col III $\alpha\beta$  - carries the human Col III gene with  $\alpha$ -MFSS and both subunits of the human Prolyl 4- hydroxylase.

*P. pastoris* nCol III - is similar to *P. pastoris* nCol III  $\alpha\beta$ , but uses the native Col III signal sequence.

25 *P. pastoris*  $\alpha\beta$  - carries both subunits of human prolyl 4-hydroxylase.

*P. pastoris*  $\alpha\beta\beta$  contains human prolyl 4-hydroxylase, wherein the  $\alpha:\beta$  gene ratio is 1:2.

*P. pastoris*  $\alpha$  contains the human prolyl 4-hydroxylase  $\alpha$  gene.

30 *P. pastoris*  $\beta$  contains the human prolyl 4-hydroxylase  $\beta$  gene.

The *P. pastoris* strains described in paragraph 5 were grown in rotary shakers to an OD600 of 5.0. Samples were taken and run on PAGE gels. Western blots were performed and analyzed with antibodies against proCol III N- terminal peptide, the  $\alpha$ -subunit of human prolyl 4-hydroxylase and the  $\beta$ -subunit of human prolyl  
35 4-hydroxylase.



The Western blots described in paragraph 6 demonstrated that both human collagen III and human prolyl 4- hydroxylase were produced in *P. pastoris*.

Pepsin digestion experiments were performed to test for triple helical structure in the human collagen produced in *P. pastoris*. Whereas most proteins are degraded by the proteolytic enzyme pepsin, the triple helical region of collagen is pepsin resistant. The collagen from cell lysates of *P. pastoris* Col III $\alpha$  $\beta$  were digested with pepsin, and the digestion products were separated by SDS-PAGE. The results of these experiments indicated that triple helical human collagen III was produced in the recombinant *P. pastoris* cells.

Experiments were performed to measure human prolyl 4-hydroxylase activity in the *P. pastoris* strains described above. *P. pastoris* has no intrinsic prolyl 4-hydroxylase activity. The assay were performed with <sup>14</sup>C labelled proline, essentially as described by Kivirikko in Methods in Enzymology, Volume 82, pgs. 245-304, Academic Press, San Diego, CA. Prolyl 4-hydroxylase activity was found in the recombinant cells.

**B. Example 12: Expression of Recombinant Collagen Genes in Mammalian Cells Expressing Recombinant Genes for Prolyl 4-Hydroxylase**

**1. Construction of a Recombinant Semliki Forest Virus Vectors Containing Collagen Genes.**

pSFVmoXIII: The Semliki Forest expression vector was constructed using the vector pBSmoXIII generated based on clones and sequences as described for pVLmoXIII above (Rehn *et al.*, submitted; Peltonen *et al.*, submitted) and the eukaryotic expression vector pSFV-1 (Liljestrom *et al.*, Bio/tecnology 2:1356-1361 (1991)). pBSmoXIII is digested with *Eco*RI to generate the full-length type XIII collagen variant with seven bp 5' untranslated region and 288 bp 3' untranslated region, and this fragment is made blunt ended with Klenow, and cloned into the *Sma*I site of pSFV-1 to give the plasmid pSFVmoXIII. pSFVmoXIII plasmid was used to produce RNA by *in vitro* transcription using MEGAscript<sup>®</sup> *in vitro* transcription kit by Ambion. Baby hamster kidney (BNK) cells transfected with the RNA as described in Liljestrom *et al.*, Current Protocols in Molecular Biology

2:16-20 (1991). Synthesis of full-length chains for mouse type XIII collagen were observed in the BHK cells by Western blotting of SDS-polyacrylamide gel-fractionated cell extracts.

5 Efficient expression of other collagen genes in cells of higher eukaryotes will be based on the above- described Semliki Forest virus vector. Semliki Forest virus is preferred as the virus because it has a broad host range such that infection of the above mentioned mammalian cell lines will also be possible. More specifically, it is expected that the use of the Semliki Forest virus can be used in a wide range of  
10 hosts, as the system is not based on chromosomal integration, and therefore it will be a quick way of obtaining modifications of the recombinant collagens in studies aiming at identifying structure-function relationships and testing the effects of various hybrid molecules. In addition, it is expected that use of the Semliki Forest virus will  
15 yield very high recombinant expression levels, over 10 ug/1x10<sup>6</sup> cells.

HeLa cells and the vaccinia virus-based expression system can also be used to express collagens in mammalian cells, and will preferably be used to expresst type IV collagens as homo- and hetero- trimer isoforms of the six type IV collagen  
20 chains.

All patents, patents applications, and publications cited are incorporated herein by reference.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. Indeed, various modifications of the  
25 above-described makes for carrying out the invention which are obvious to those skilled in the field of immunology, biochemistry, or related fields are intended to be within the scope of the following claims.

30

35

**CLAIMS****WHAT IS CLAIMED IS:**

1. A method for producing a collagen polypeptide, wherein said collagen  
5 is selected from the group comprising collagen types IV, V, VI, VII, VIII, IX, X,  
XI, XII, XIII, XIV, XV, XVI, XVII, XVIII, and XIX, comprising:
  - a. culturing a host cell, wherein said host cell has been infected,  
transfected or transformed with (i) a first expression vector comprising a  
polynucleotide molecule having a nucleic acid sequence which encodes a collagen  
10 subunit; and (ii) a second expression vector comprising a polynucleotide molecule  
having a nucleic acid sequence which encodes at least one collagen post-translational  
enzyme or subunit thereof; and
  - b. purifying said collagen polypeptide.
- 15 2. The method of Claim 1 wherein the host cell is selected from the  
group consisting of a yeast cell, a plant cell, an insect cell and a mammalian cell.
- 20 3. The method of Claim 1 wherein the host cell is further infected,  
transfected or transformed with a third expression vector comprising a polynucleotide  
molecule having a nucleic acid sequence which encodes a second collagen subunit.
- 25 4. The method of Claim 3 wherein the host cell is further infected,  
transfected or transformed with a fourth expression vector comprising a  
polynucleotide molecule having a nucleic acid sequence which encodes a third  
collagen subunit.
- 30 5. The method of Claim 1 wherein said collagen post-translational  
enzyme is selected from the group consisting of prolyl-4-hydroxylase, lysyl oxidase,  
lysyl hydroxylase, C-proteinase, and N-proteinase.

35

6. The method of Claim 1 wherein the collagen post-translational enzyme subunit is selected from the group consisting of an alpha subunit of prolyl-4-hydroxylase and a beta subunit of prolyl-4-hydroxylase.

5

7. A method for producing a procollagen polypeptide, wherein said procollagen is selected from the group comprising collagen types IV, V, VI, VII, VIII, IX, X, XI, XII, XIII, XIV, XV, XVI, XVII, XVIII, and XIX, comprising:

- a. culturing a host cell, wherein said host cell has been infected,  
10 transfected or transformed with: (i) a first expression vector comprising a polynucleotide molecule having a nucleic acid sequence which encodes a collagen subunit; and (ii) a second expression vector comprising a polynucleotide molecule having a nucleic acid sequence which encodes at least one collagen post-translational  
15 enzyme or subunit thereof; and
- b. purifying said procollagen polypeptide.

8. The method of Claim 7 wherein the host cell is selected from the  
20 group consisting of a yeast cell, a plant cell, an insect cell and a mammalian cell.

9. The method of Claim 7 wherein the host cell is further infected,  
transfected or transformed with a third expression vector comprising a polynucleotide  
25 molecule having a nucleic acid sequence which encodes a second collagen subunit.

10. The method of Claim 9 wherein the host cell is further infected,  
transfected or transformed with a fourth expression vector comprising a  
polynucleotide molecule having a nucleic acid sequence which encodes a third  
30 collagen subunit.

11. The method of Claim 7 wherein said collagen post-translational  
enzyme is selected from the group consisting of prolyl-4-hydroxylase, lysyl oxidase  
35 and lysyl hydroxylase.

12. The method of Claim 7 wherein the collagen post-translational enzyme subunit is selected from the group consisting of an alpha subunit of prolyl-4-hydroxylase and a beta subunit of prolyl-4-hydroxylase.

5

13. A collagen polypeptide, wherein said collagen is selected from the group comprising collagen types IV, V, VI, VII, VIII, IX, X, XI, XII, XIII, XIV, XV, XVI, XVII, XVIII, and XIX, manufactured according to a method comprising:

a. culturing a host cell, wherein said host cell has been  
10 infected, transfected or transformed with: (i) a first expression vector comprising a polynucleotide molecule having a nucleic acid sequence which encodes a collagen subunit; and (ii) a second expression vector comprising a polynucleotide molecule having a nucleic acid sequence which encodes at least one collagen post-translational  
15 enzyme or subunit thereof; and

b. purifying said collagen polypeptide.

14. The collagen polypeptide of Claim 13 wherein the host cell is selected  
20 from the group consisting of a yeast cell, a plant cell, an insect cell and a mammalian cell.

15. The collagen polypeptide of Claim 13 wherein the host cell is further  
25 infected, transfected or transformed with a third expression vector comprising a polynucleotide molecule having a nucleic acid sequence which encodes a second collagen subunit.

16. The collagen polypeptide of Claim 15 wherein the host cell is further  
30 infected, transfected or transformed with a fourth expression vector comprising a polynucleotide molecule having a nucleic acid sequence which encodes a third collagen subunit.

35 17. The collagen polypeptide of Claim 13 wherein said collagen post-translational enzyme is selected from the group consisting of

prolyl-4-hydroxylase, lysyl oxidase, lysyl hydroxylase, C-proteinase, and N-proteinase.

18. The collagen polypeptide of Claim 13 wherein the collagen  
5 post-translational enzyme subunit is selected from the group consisting of an alpha  
subunit of prolyl-4-hydroxylase and a beta subunit of prolyl-4- hydroxylase.

19. The collagen polypeptide of Claim 13 wherein said polypeptide is not  
glycosolated.  
10

20. The collagen polypeptide of Claim 13 wherein said polypeptide is  
partially deglycosolated.

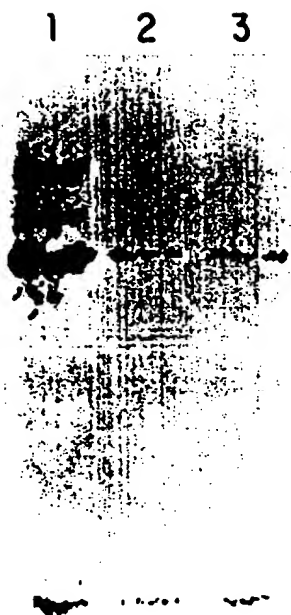
15 21. A host cell which has been infected, transfected or transformed with:  
(i) a first expression vector comprising a polynucleotide molecule having a nucleic  
acid sequence which encodes a collagen subunit; and (ii) a second expression vector  
comprising a polynucleotide molecule having a nucleic acid sequence which encodes  
20 at least one collagen post-translational enzyme or subunit thereof.

22. The host cell of Claim 21 wherein said host cell is further infected,  
transfected or transformed with a third expression vector comprising a second  
collagen subunit.  
25

23. The host cell of Claim 22 wherein said host cell is further infected,  
transfected or transformed with a fourth expression vector comprising a third  
collagen subunit.  
30

35

1/10



**FIG. 1**

2/10

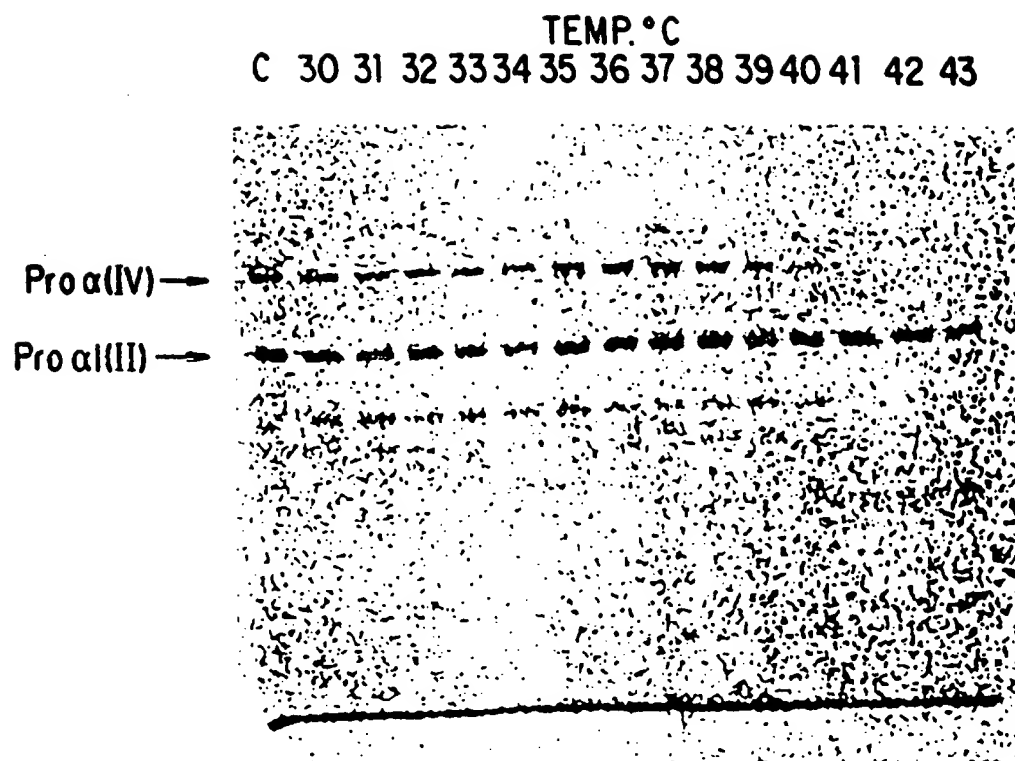


FIG. 2



3/10

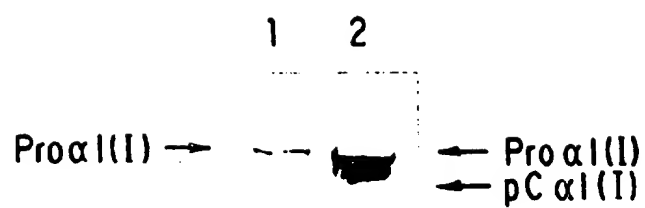


FIG. 3

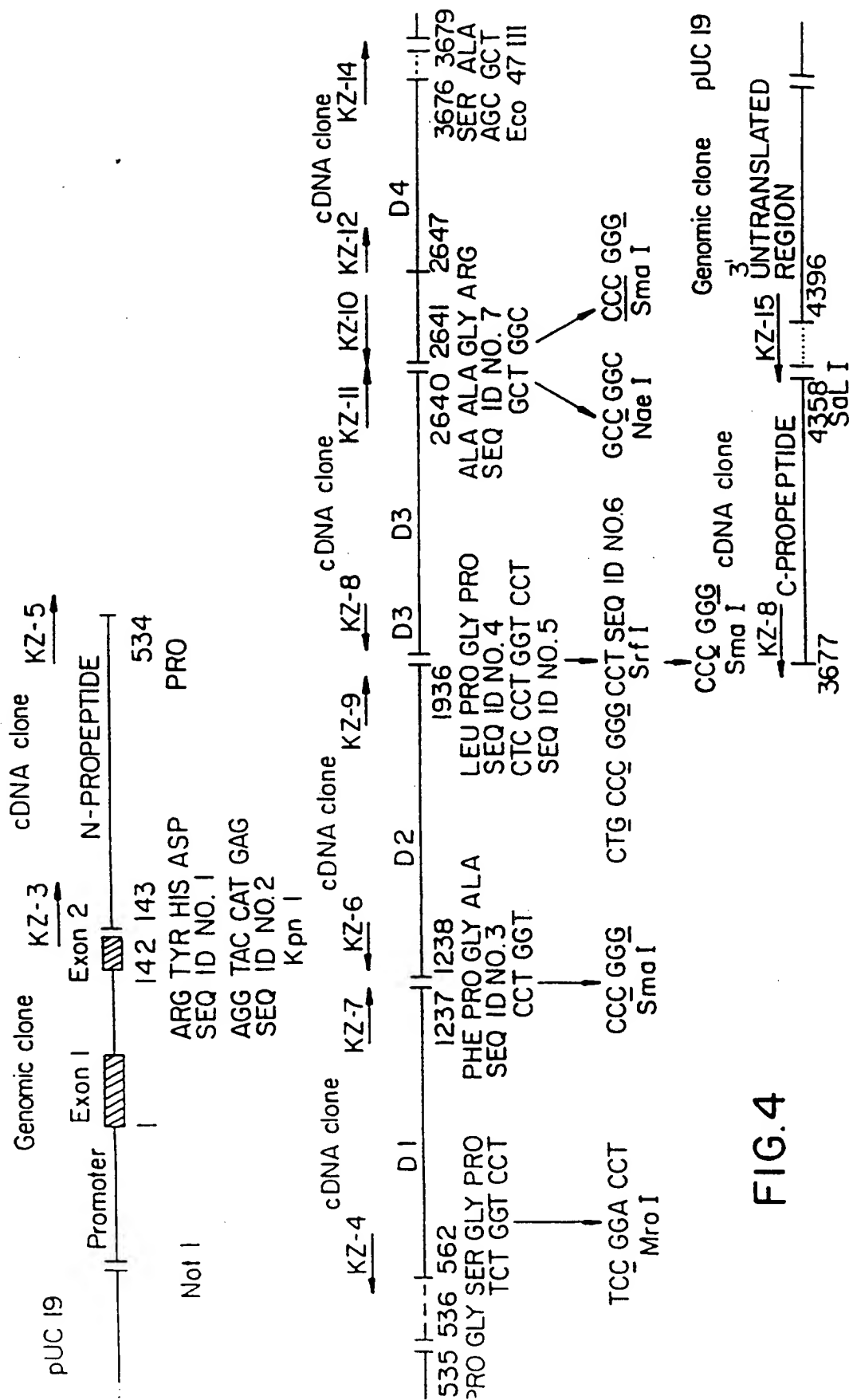


FIG. 4

5/10



FIG. 5

6/10

FIG. 6A

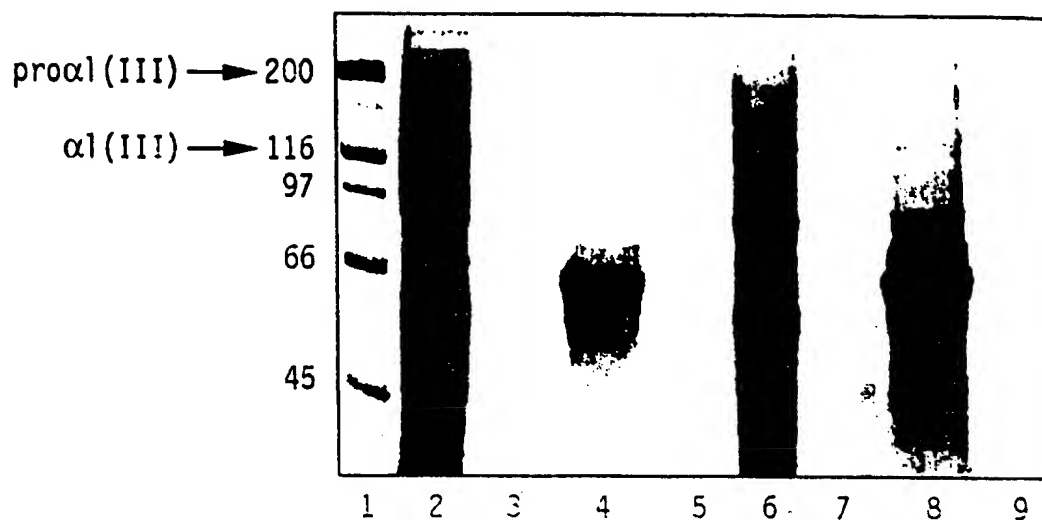
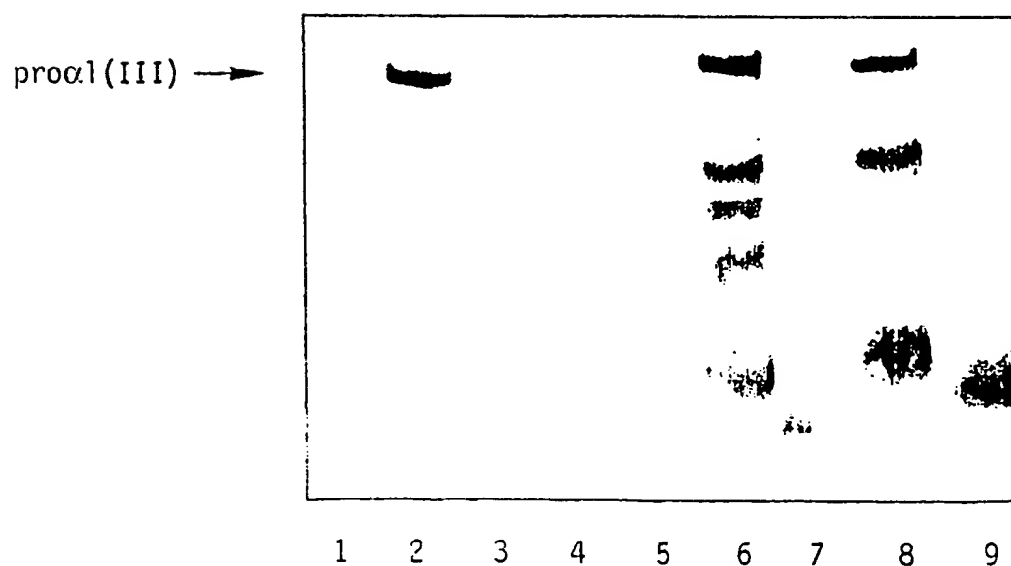


FIG. 6B



7/10

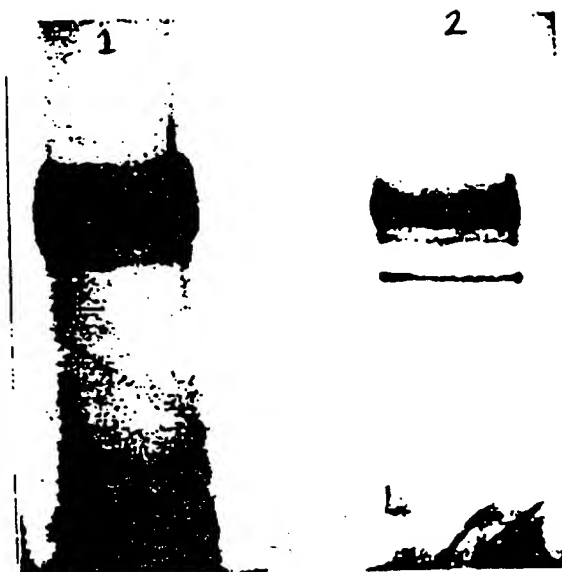


FIG. 7

8/10

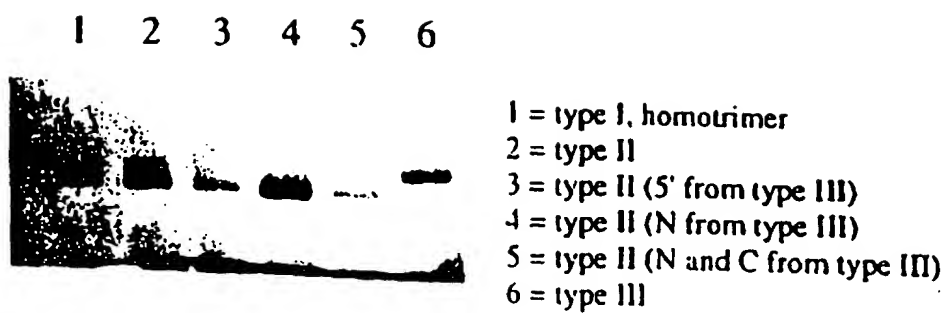


FIG. 8

FIG. 9

9/10

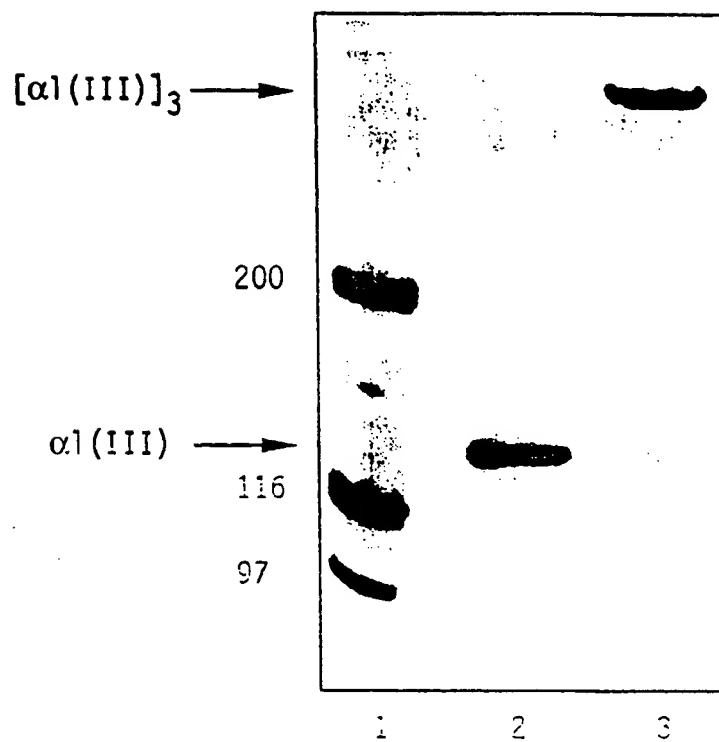


FIG. 10

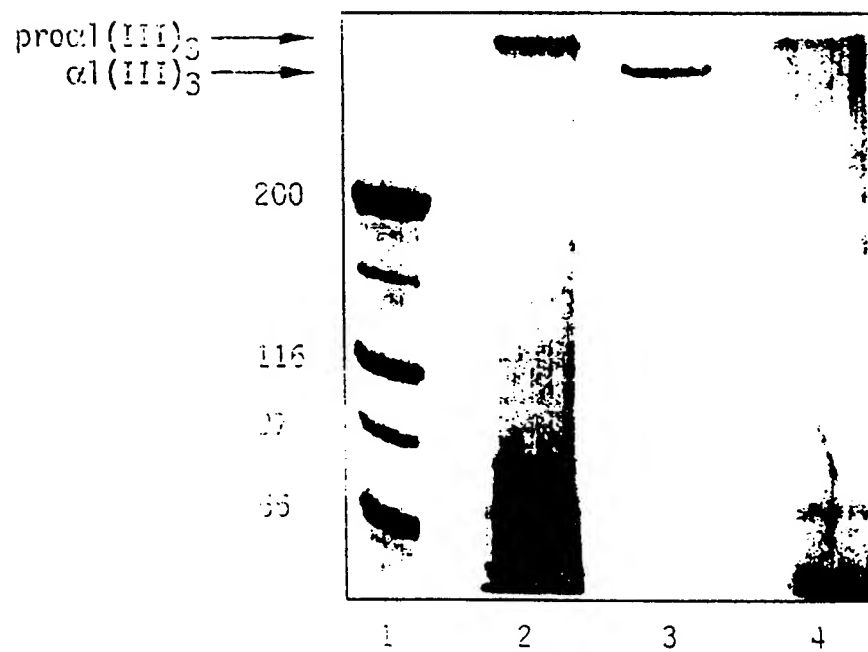


FIG. 11A

10/10

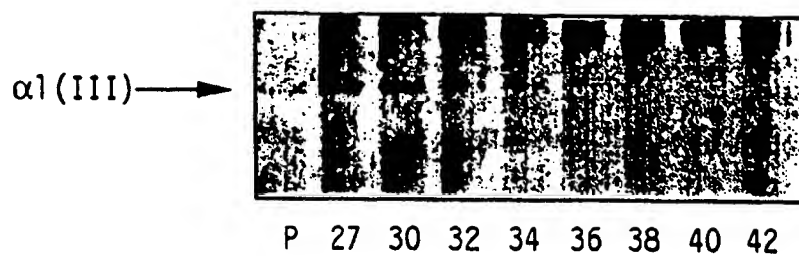


FIG. 11B

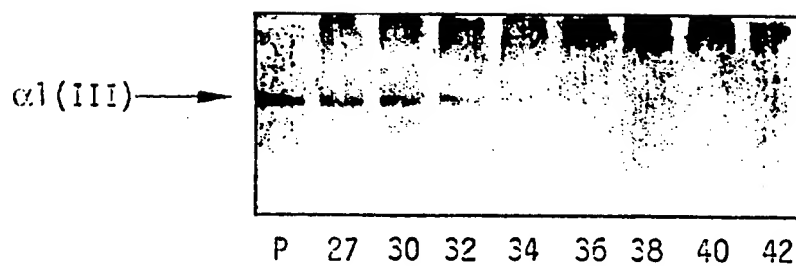


FIG. 11C

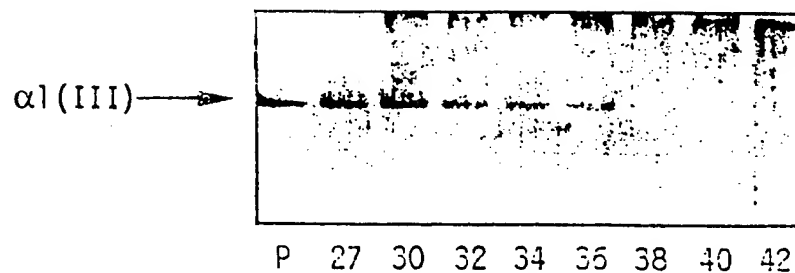
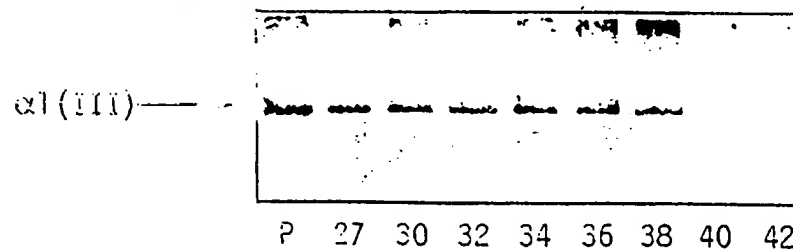


FIG. 11D





## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/07300

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) : A61K 38/39, 38/17; C12P 21/00

US CL : 530/356; 435/69.1, 252.3

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/356; 435/69.1, 252.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Dialog, APS

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 93/07889 A1 (THOMAS JEFFERSON UNIVERSITY) 29 April 1993 (29.04.93), see entire document.	1-23
Y	HELAAKOSKI et al. Structure and Expression of the Human Gene for the $\alpha$ Subunit of Prolyl 4-Hydroxylase. J. Biol. Chem. 11 November 1994 Vol. 269, No. 45, pages 27847-27854, see entire document.	1-23
Y	ADAMS, S.L. Collagen Gene Expression. Amer. J. of Respiratory Cell and Molecular Biology. 1989, Vol. 1 pages 161-168, see entire document.	1-23
Y	RYAN et al. The Human Type II Procollagen Gene: Identification of an Additional Protein-Coding Domain and Location of Potential Regulatory Sequences in the Promoter and First Intron. Genomics. 1990, Vol. 8, pages 41-48, see entire document.	1-23

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:	* T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
* A* document defining the general state of the art which is not considered to be of particular relevance	* X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
* E* earlier document published on or after the international filing date	* Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
* L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	* &*	document member of the same patent family
* O* document referring to an oral disclosure, use, exhibition or other means		
* P* document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

28 JULY 1997

Date of mailing of the international search report

20 AUG 1997

 Name and mailing address of the ISA/US  
 Commissioner of Patents and Trademarks  
 Box PCT  
 Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

ENRIQUE D. LONGTON

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/07300

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MAYNE et al. New Members of the Collagen Superfamily. Current Opinion in Cell Biology. 1993, Vol. 5, pages 883-890, see entire document.	1-23
Y,P	US 5,593,859 A (PROCKOP et al.) 14 January 1997, see entire document.	1-23
Y	US 5,405,757 A (PROCKOP et al.) 11 APRIL 1995, see entire document.	1-23